Epipolar-Aligned Channel Selection: A Projection from Optical Flow to Disparity

Sahereh Obeidavi
Coburg University of Applied Science
Faculty of Electrical Engineering and Computer Science
Coburg, Germany

Dieter Landes
Coburg University of Applied Science
Faculty of Electrical Engineering and Computer Science
Coburg, Germany

Arsalan Moosavipoor
Islamic Azad University, Central Tehran Branch
Department of Electrical Power Engineering

ABSTRACT

Stereo disparity estimation is a fundamental problem in computer vision, forming the basis for 3D reconstruction, autonomous navigation, and robotics. Unlike optical flow, which describes unconstrained 2D displacements, disparity in rectified stereo geometry is strictly aligned with the epipolar axis. This geometric property implies that one component of the flow field contains the true disparity signal, while the orthogonal component predominantly reflects distortion, miscalibration, or noise. However, most existing approaches either neglect this constraint or require dedicated disparity networks trained from scratch, leading to redundant computation and limited generality.

This paper introduces Epipolar-Aligned Channel Selection (EACS), a parameter-free and geometry-aware post-processing operator that isolates the disparity-aligned component of optical flow while discarding the non-epipolar channel. Implemented as a fixed linear projection with negligible overhead, EACS ensures that only geometrically meaningful information is retained. When coupled with RAFT, a state-of-the-art optical flow network, the resulting RAFT + EACS pipeline enables direct and efficient disparity estimation from optical flow, without requiring additional training or specialized stereo architectures.

Experiments conducted on synthetic stereo data generated at TU Chemnitz (Technische Universität Chemnitz) confirm the effectiveness of this approach. The proposed method achieves sub-pixel disparity accuracy (MAE = 0.3007, RMSE = 0.9470) and extremely low error rates under stringent evaluation protocols (D1-all = 0.4%). Qualitative analysis further demonstrates that RAFT + EACS preserves fine structural details and produces smooth, consistent disparity maps, even in challenging low-texture regions. These findings establish geometry-aware post-processing as a simple yet powerful alternative to specialized stereo disparity networks.

General Terms

Computer Vision, Image Processing, 3D Reconstruction, Algorithms, Performance Evaluation.

Keywords

Stereo disparity estimation, Optical Flow, Epipolar Geometry, RAFT, Epipolar-Aligned Channel, Optical Flow–to–Disparity

1. INTRODUCTION

The estimation of 3D motion from visual data has long been a central challenge in computer vision, with applications spanning autonomous navigation, robotics, immersive media, and environmental monitoring. A key formulation of this problem is *scene flow*, which represents the dense 3D motion field of points in a scene [1]. Traditionally, scene flow can be decomposed into two tightly related sub-tasks: *optical flow*, capturing 2D displacements between temporally adjacent frames [2–4], and *stereo disparity*, describing pixel correspondences across left–right stereo pairs [5–8]. By combining these two complementary modalities, depth and motion can be jointly inferred, enabling a full reconstruction of dynamic 3D geometry.

Despite their conceptual similarity, optical flow and stereo disparity are often treated as distinct problems, each with its own datasets, architectures, and optimization objectives. Optical flow estimation typically searches for correspondences over the entire image domain [2, 4, 7], whereas stereo disparity estimation restricts matching to epipolar lines determined by the stereo baseline [5, 6, 8]. This geometric distinction has motivated separate model designs and training pipelines. However, such separation can lead to inefficiencies: information that is useful for one task (e.g., flow smoothness priors, stereo consistency) is not fully exploited by the other. Recent advances in joint modeling, building on high-quality optical flow backbones such as RAFT [9] and RAFT-Stereo [10], have shown that shared architectures can effectively leverage cross-task regularities across optical flow, stereo, and depth. More recent transformer-based joint frameworks further unify pose, depth, and optical flow within a single architecture, underscoring the benefit of exploiting geometric relationships across tasks [11].

In rectified stereo geometry, only one component of the optical flow field is geometrically meaningful for disparity estimation. The two-channel flow vector (f_x, f_y) captures apparent 2D motion between two views, but in left–right stereo setups the true disparity signal lies almost entirely along the horizontal axis, while the vertical component contains only distortion, calibration residuals, or wide-FOV artefacts. Conversely, in top–bottom rigs the vertical component carries the geometry, and the horizontal component becomes negligible. Under ideal epipolar geometry, the disparity is strictly constrained to the baseline direction and the orthogonal component should theoretically vanish. These structural properties motivate the central idea of this work: isolating the epipolar-aligned flow component provides exactly the information required for disparity estimation, while discarding the orthogonal component removes nuisance variation that is irrelevant to the task.

A critical yet underexplored issue in this joint setting is the *representation gap* between optical flow and disparity. Raw optical flow fields contain both horizontal (f_x) and vertical (f_y) displacement components, while disparity in rectified stereo geometry is constrained to a single axis aligned with the baseline (horizontal for conventional rigs, vertical for top–

bottom stereo). As a result, one of the flow channels predominantly carries the true disparity signal, whereas the orthogonal channel mainly reflects distortions, miscalibration, or noise. Recent depth-estimation approaches that explicitly leverage optical flow as an auxiliary supervisory signal further demonstrate the usefulness of flow–geometry interactions for depth prediction [12]. Feeding the full 2D flow into disparity-related networks introduces redundancy and may even degrade performance.

In this work, this gap has been addressed by proposing Epipolar-Aligned Channel Selection (EACS), a lightweight and differentiable post-processing operator that is applied after optical flow estimation to extract the disparity-aligned component while suppressing the orthogonal component. This is a challenge because the orthogonal component often has large magnitude but carries no geometric meaning; without explicitly removing it, downstream networks must implicitly learn to ignore this misleading signal, which increases training complexity and leads to unstable or suboptimal disparity predictions. . Implemented as a fixed 1×1 convolution, EACS introduces no trainable parameters and negligible computational overhead, yet it ensures that only the geometrically meaningful signal is propagated downstream. This formulation offers both theoretical grounding—since it directly encodes epipolar constraints—and practical benefits by reducing nuisance variation. By integrating EACS into existing optical-flow–to–disparity pipelines as a post-processing projection on the flow output (e.g., from RAFT), this work shows that stereo disparity estimation can be made more robust and efficient. Moreover, our approach is fully compatible with modern architectures such as RAFT [9] and RAFT-Stereo [10], and can be seamlessly deployed in joint optical flow and stereo disparity networks.

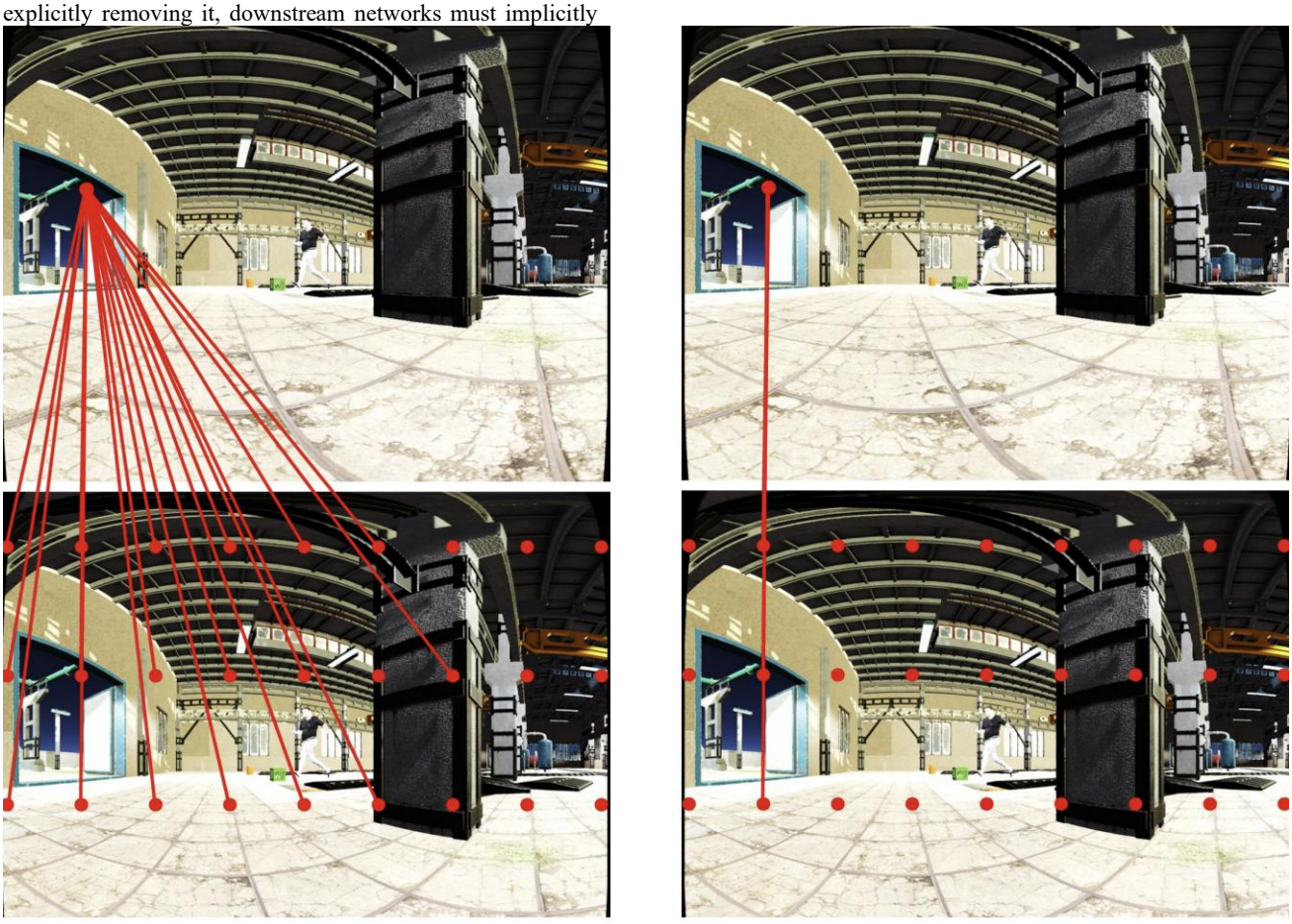

Fig 1: Conceptual relationship between optical flow and stereo disparity estimation. (a): optical flow estimation involves dense correspondence search across the entire two-dimensional image plane. (b): stereo disparity estimation restricts matching to the epipolar line, reducing the search space to a single axis.

To further motivate our approach, Fig 1 highlights the conceptual relationship between optical flow and stereo disparity. Both tasks can be understood as dense correspondence problems: optical flow seeks pixel matches across temporally adjacent frames with a two-dimensional search space, while stereo disparity restricts the matching problem to the one-dimensional epipolar line determined by the stereo baseline. This structural similarity suggests that disparity can, in principle, be recovered directly from optical flow if the search space is appropriately constrained. This motivates using optical flow as a surrogate representation for disparity estimation and provides the conceptual foundation for the proposed EACS operator, which explicitly enforces epipolar alignment. Recent generative approaches have also begun to explore depth estimation through flow-based transformations, reinforcing the relevance of flow–depth relationships even beyond discriminative frameworks [13].

The contributions of this work are both methodological and conceptual. At the methodological level, this paper introduces *Epipolar-Aligned Channel Selection (EACS)* as a geometry-aware post-processing operator that enforces epipolar

constraints by projecting dense optical flow fields onto the stereo baseline. This simple yet principled design eliminates the influence of the non-epipolar channel while preserving the disparity-aligned signal, and it can be implemented as a parameter-free, differentiable 1×1 convolution. At the conceptual level, this work reframes disparity estimation not as an independent learning problem requiring a dedicated network, but as a constrained projection of optical flow, highlighting the sufficiency of geometry-aware post-processing in bridging the two tasks. Finally, through systematic experiments on synthetic stereo data with dense ground truth, this paper demonstrates that RAFT + EACS achieves sub-pixel disparity accuracy while preserving fine structural detail, all with negligible computational overhead. These findings underscore the broader significance of incorporating simple, theoretically grounded operators into deep pipelines, showing that lightweight geometry-aware post-processing can serve as an effective alternative to dedicated disparity estimation networks.

2. RELATED WORK

2.1 Stereo Disparity Estimation

Stereo disparity estimation has long been recognized as a fundamental component of 3D reconstruction. Classical approaches typically relied on local correlation windows or global energy minimization frameworks, often incorporating smoothness priors and occlusion handling [2, 3, 14]. While effective under controlled conditions, these methods were prone to failure in the presence of noise, illumination changes, or textureless regions, reflecting the limitations of handcrafted optimization schemes.

With the rise of deep learning, convolutional neural networks (CNNs) have been applied successfully to optical flow and stereo disparity estimation, achieving superior accuracy and performance [5, 7]. Zbontar et al. [15] first used CNNs to learn image patch similarities, inspiring subsequent encoder–decoder architectures [16–18]. The introduction of end-to-end stereo networks, such as PSMNet [6] and GC-Net [5], marked a turning point by directly regressing disparity from rectified stereo pairs. Central to these methods is the construction of cost volumes, either in 3D through correlations between left–right features [16] or in 4D by concatenating features to preserve channel dimensions. Architectures adopting 3D or 4D convolutions for cost aggregation [19, 20] have proven highly successful, particularly when integrating classical concepts such as semi-global matching [19]. Despite these advances, stereo disparity remains inherently constrained by epipolar geometry: disparities must align with the stereo baseline, a property that deep networks do not always explicitly encode.

2.2 Optical Flow Estimation

Optical flow estimation generalizes disparity prediction to arbitrary temporal displacements in video. CNN-based models such as FlowNet [17] and PWC-Net [7] significantly advanced the field by introducing encoder–decoder pipelines, warping mechanisms, and cost-volume operations in multiple resolutions [18, 21]. While cost-volume–based approaches improved accuracy, 4D convolutions were computationally and memory intensive, often requiring millions of iterations to train.

To address this, the RAFT architecture [9] introduced an iterative refinement strategy using a gated recurrent unit (GRU) with high-resolution correlation volumes. RAFT reduced model size while improving accuracy on standard benchmarks, and it remains a cornerstone for both optical flow and scene-flow estimation [22–24]. More recently, RAFT-Stereo [10] extended this framework to disparity estimation by aligning correlation volumes with the epipolar direction.

Beyond classical CNN-based flow models, several recent works have highlighted the strong coupling between optical flow and geometric structure. F²Depth [12] employs optical-flow consistency and feature-map synthesis losses to supervise self-supervised monocular depth estimation, showing that accurate flow is an effective geometric supervisory signal. In parallel, DepthFM [13] formulates monocular depth estimation as a flow-matching transport problem, demonstrating that trajectory-based flow modeling can enhance both training and inference efficiency. These developments further emphasize the close relationship between optical flow and geometric quantities—an observation that directly motivates geometry-aware refinements of flow for disparity estimation.

Nevertheless, raw optical flow inherently contains both horizontal and vertical components, which are not equally meaningful in stereo setups where disparity is restricted to a single axis. This discrepancy between general-purpose flow and epipolar geometry motivates specialized refinements.

2.3 Multi-Task and Joint Models

Given the close relationship between stereo disparity and optical flow, a variety of works have sought to unify them. Early efforts were inspired by variational methods that applied similar objectives to both tasks [25, 26]. Neural-network–based approaches later demonstrated that sharing encoders or correlation volumes can improve both disparity and flow estimation by leveraging cross-task regularities [27, 28]. Beyond pairwise matching, multi-task learning has been extended to broader scene-flow estimation. For example, DispNet and FlowNet were combined with occlusion estimation for joint scene-flow prediction [29], while PWC-Net [7] variants integrated stereo, flow, and semantic segmentation within a shared encoder [30]. RAFT-3D [23] further advanced this line by combining RAFT's recurrent refinement with pre-estimated depth to predict full 3D motion under rigid-motion constraints. Transformer-based approaches [31] have also shown promise across both tasks.

More recently, transformer-driven joint frameworks have explored even tighter geometric coupling. PDF-Former [11] jointly estimates pose, depth, and optical flow through a competition–cooperation mechanism, demonstrating that transformer architectures can effectively exploit shared structure across geometric tasks and benefit from mutual supervision. These modern multi-task approaches highlight the potential of unified representations, yet most still treat disparity and flow as distinct outputs requiring dedicated network heads. In doing so, they fail to exploit the theoretical fact that stereo disparity is not an independent modality but a projection of optical flow along the epipolar axis—resulting in unnecessary architectural complexity and redundancy.

2.4 Summary and Motivation

Existing stereo and optical-flow methods have made substantial progress, yet they typically treat the two tasks as separate problems with independent network branches and training objectives. Despite the conceptual overlap between them, current approaches rarely exploit the fact that, in rectified stereo geometry, the disparity signal corresponds to a single epipolar-aligned component of the optical flow field. As a result, most models preserve and process both flow channels, introducing redundancy and additional learning complexity.

This observation gives rise to the central objective of this work: to determine whether accurate stereo disparity can be recovered directly from optical flow by isolating only the epipolar-aligned component and discarding the orthogonal one, without relying on a dedicated stereo network.

To investigate this objective, this work proposes Epipolar-Aligned Channel Selection (EACS), a lightweight post-processing operator that projects dense optical flow onto the baseline direction, removing the non-epipolar component while preserving the disparity-relevant signal. The method is parameter-free, compatible with modern flow architectures, and designed as a minimal test of the theoretical sufficiency of the epipolar-aligned optical-flow component.

3. METHODOLOGY

As discussed in Section 1, in rectified stereo geometry only one component of the optical flow field is aligned with the epipolar direction and therefore carries the disparity-relevant signal, while the orthogonal component contains residual distortions or wide-FOV artefacts. Building on this observation, our goal is to construct an operator that isolates the epipolar-aligned flow component and suppresses the non-epipolar one in a principled and computationally lightweight manner.

3.1 Problem Formulation

The objective of this work is to recover a single-channel stereo disparity map directly from a two-channel optical flow field, without training a stereo network or modifying the underlying flow architecture. In optical flow representation, each pixel (i,j) is associated with a displacement vector consisting of horizontal and vertical components. Formally, the flow field can be written as:

$$f(i,j) = \begin{bmatrix} f_x(i,j) \\ f_y(i,j) \end{bmatrix} \quad (1)$$

where $f_x(i,j)$ denotes the displacement along the horizontal axis and $f_y(i,j)$ denotes the displacement along the vertical axis. Collecting these vectors over the entire image yields the flow tensor:

$$F \in \mathbb{R}^{2\times H\times W} \quad (2)$$

where H, W are the spatial dimensions.

Our goal is to construct a mapping that discards the horizontal channel and preserves only the vertical channel. This mapping can be expressed as:

$$\mathcal{M}: \mathbb{R}^{2\times H\times W} \rightarrow \mathbb{R}^{1\times H\times W} \quad (3)$$

$$\mathcal{M}(F) = e_x F_x + e_y F_y \quad (4)$$

Where for vertical baseline $e = [0,1]^T$, this reduces to F_y, and for horizontal baseline $e = [1,0]^T$, this reduces to F_x.

To implement this mapping in the form of a neural network, a single convolutional layer with kernel size 1×1 is used. The convolution operation at each spatial location is defined as:

$$y(i,j) = w_x . f_x(i,j) + w_y . f_y(i,j) \quad (5)$$

where w_x and w_y are the weights associated with the two input channels. In order to select only the vertical channel, by setting:

$$w_x = 0, \qquad w_y = 1 \quad (6)$$

which reduces (5) to:

$$y(i,j) = f_y(i,j) \quad (7)$$

This formulation ensures that the output is exactly the vertical flow component, while the horizontal component is completely suppressed.

By expressing the channel selection as a fixed 1×1 convolution, the operation remains differentiable and compatible with common deep learning frameworks. The network contains no trainable parameters, which guarantees negligible computational overhead, yet it can be exported and deployed as part of larger models without requiring special handling. In this way, the simple task of discarding the x-channel is achieved in a mathematically principled and framework-friendly manner. the operation remains differentiable and compatible with common deep learning frameworks. The network contains no trainable parameters, which guarantees negligible computational overhead, yet it can be exported and deployed as part of larger models without requiring special handling. In this way, the simple task of discarding the x-channel is achieved in a mathematically principled and framework-friendly manner.

3.2 Epipolar-Aligned Channel Selection (EACS)

To generalize the fixed channel selection described in Section 3.1, the Epipolar-Aligned Channel Selection (EACS) operator is formalized as a projection onto the stereo baseline direction. Given an optical flow vector $f(i,j) = [f_x(i,j),\ f_y(i,j)]^T$, a baseline unit vector e is defined as $e = \frac{b}{\|b\|}$, where b is the stereo baseline vector. For rectified horizontal rigs $e = [1,0]^T$ (retain f_x); and for vertical rigs $e = [0,1]^T$ (retain f_y).

EACS computes a projection of the two-channel flow onto the (unit) baseline direction e:

$$f_{EACS}(i,j) = e^T . f(i,j) = e_x . f_x(i,j) + e_y . f_y(i,j) \quad (8)$$

This operation discards the orthogonal component and retains only the disparity-aligned signal. In implementation, Eq. (8) corresponds exactly to the fixed 1×1 convolution described earlier, with weights determined by the baseline orientation.

To provide an intuitive overview of the proposed operator, Fig 2 illustrates the internal structure of the Epipolar-Aligned Channel Selection (EACS) module. Starting from dense optical flow fields (f_x, f_y) of size $2 \times H \times W$, EACS applies a fixed 1×1 convolution that selects only the epipolar-aligned channel while discarding the orthogonal one. The resulting single-channel feature map directly constitutes the estimated disparity The resulting single-channel output constitutes the estimated disparity, which can later be evaluated using standard disparity metrics.

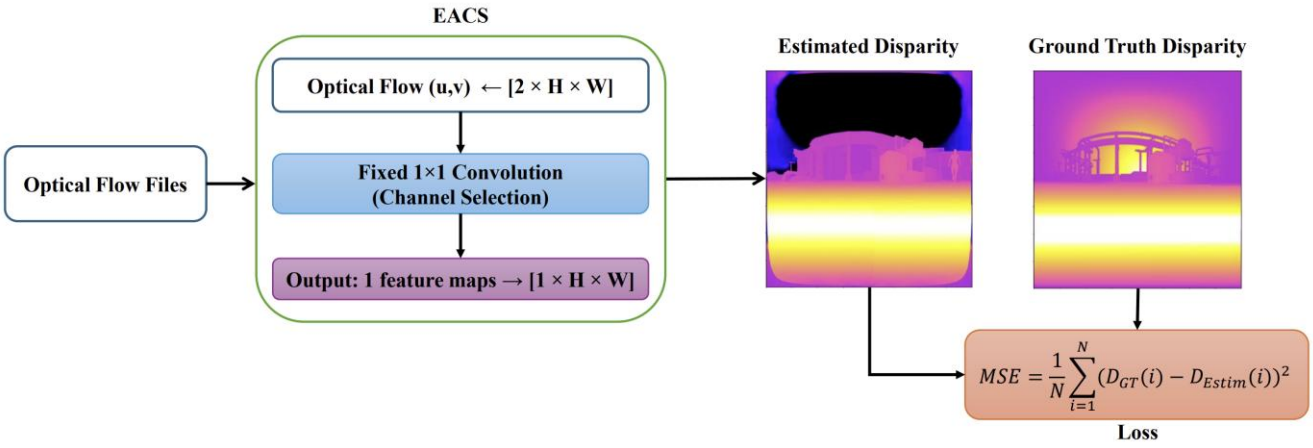

Fig 2: Schematic representation of the Epipolar-Aligned Channel Selection (EACS) operator. The module processes two-channel optical flow (u, v) through a fixed 1×1 convolution to suppress the non-epipolar component. The resulting single-channel output corresponds to the estimated disparity map, which is subsequently evaluated against ground-truth disparity using standard metrics

3.3 Integration into Existing Architectures

In our framework, disparity estimation is obtained by coupling RAFT with the proposed Epipolar-Aligned Channel Selection (EACS) module, which is straightforward and does not require retraining. The input stereo image pairs (I_l, I_r) for horizontal stereo or (I_t, I_b) for vertical stereo, are first processed by RAFT [9], which produces a dense two-channel optical flow field $F = [f_x, f_y]^T \in R^{2\times H\times W}$. The flow field is then passed through Epipolar-Aligned Channel Selection (EACS) operator which projects this field onto the epipolar axis, producing a single-channel disparity map $D = f_{EACS} \in R^{1\times H\times W}$.

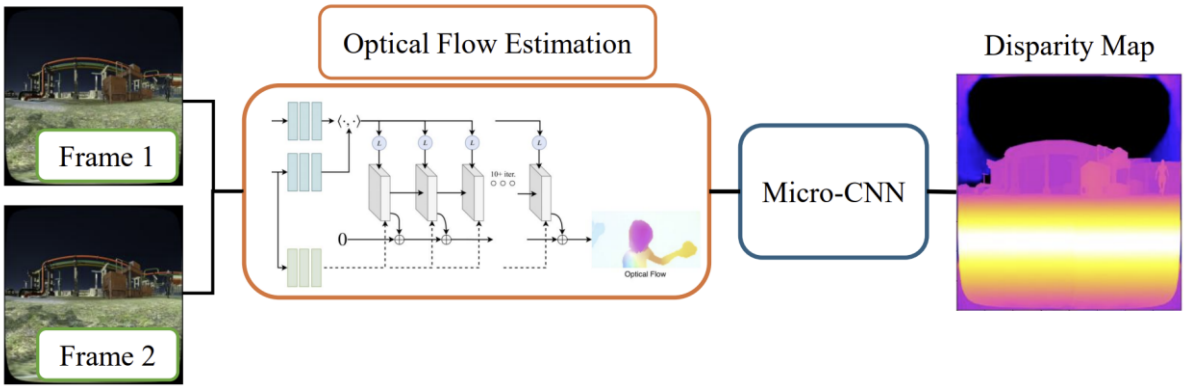

Fig 3: Overall pipeline of the proposed RAFT + EACS framework. Two rectified stereo frames are first processed by RAFT to compute dense optical flow. The flow field is then passed through the Epipolar-Aligned Channel Selection (EACS) module, implemented as a fixed 1 × 1 convolution, which removes the non-epipolar component. The output is a single-channel disparity map, obtained without the need for a dedicated stereo network.

In the case of a horizontal-baseline stereo configuration, the disparity signal is aligned with the horizontal axis, and thus EACS suppresses the vertical component f_y, retaining only the horizontal component f_x. Conversely, for a vertical-baseline stereo configuration, the disparity manifests along the vertical axis, and therefore EACS suppresses the horizontal component f_x, preserving only the vertical component f_y. In both cases, the operation yields a single-channel disparity map, that is directly consistent with the underlying geometry and can be without dedicated stereo estimation networks such as RAFT-Stereo [10] and CREStereo [32]. Fig 3 illustrates the complete RAFT + EACS pipelinein. Stereo image pairs are first processed by RAFT to generate dense two-channel optical flow, which is then passed through the proposed Epipolar-Aligned Channel Selection (EACS) operator. Implemented as a fixed 1×1 convolution, EACS discards the non-epipolar component and outputs the disparity-aligned signal. The resulting one-channel map directly constitutes the disparity estimate, demonstrating that the entire disparity estimation process can be achieved without additional trainable components.

3.4 Theoretical Analysis

The theoretical grounding of this approach can be expressed using projection matrices. Let $P = ee^T$, where e is the epipolar unit vector. In the case of horizontal stereo,

$$P_{horizontal} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (9)$$

while for vertical stereo,

$$P_{vertical} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (10)$$

Applying P to a flow vector f yields:

$$f_{EACS} = P.f \quad (11)$$

From an error propagation perspective, retaining both flow components introduces unnecessary variance into the disparity estimation process. Since disparity is strictly aligned with the epipolar direction in rectified stereo geometry, any orthogonal flow component acts as structured noise. By explicitly projecting the flow vector onto the baseline direction, EACS removes this noise at the representation level, reducing ambiguity and stabilizing downstream disparity estimation. Unlike learned suppression mechanisms, this projection is deterministic and guaranteed to preserve the physically meaningful component of motion, thereby improving robustness without increasing model complexity. Applying the projection $P = ee^T$ retains only the flow component aligned with the baseline direction, which is the quantity relevant for disparity estimation.In practical terms, the operator reduces the dimensionality of the flow representation from $2 \times H \times W$ to $1 \times H \times W$, lowering the memory footprint and simplifying subsequent cost-volume construction. Because it is implemented as a convolutional layer with fixed weights, EACS can be exported in standard formats such as ONNX or TorchScript, and integrated into real-time pipelines without any modification to the backbone network. This makes the approach highly practical for deployment in resource-constrained environments.

4. EVALUATION

The goal of our experiments is to evaluate the central hypothesis that accurate stereo disparity can be recovered directly from optical flow by projecting the flow onto the epipolar (baseline) direction and suppressing the orthogonal component. In the proposed pipeline, rectified stereo pairs are first processed by RAFT to produce dense optical flow; the resulting flow field is then passed through the Epipolar-Aligned Channel Selection (EACS) operator, which performs the projection and yields a single-channel disparity map. This procedure introduces no trainable parameters and incurs negligible computational overhead.

Accuracy is reported using MAE and RMSE (pixels) and follow the KITTI D1-all protocol [33] (error >3 px and >5% of ground truth). Threshold outlier rates (>3 px, >5 px) and runtime measured on a system equipped with an NVIDIA TU102-based GPU is also provided. While the current evaluation focuses on a controlled synthetic dataset, this choice was made deliberately to isolate the geometric effect of the proposed EACS operator under ideal calibration conditions. By eliminating confounding factors such as sensor noise, rolling shutter effects, and imperfect rectification, the experiments provide a clear cause–effect validation of the central hypothesis. This controlled setting allows us to rigorously assess whether disparity can be recovered from optical flow through epipolar-aligned projection alone. The remainder of this section details the setup (section 4.1), quantitative results (section 4.2), qualitative analysis (section 4.3), and discussion/limitations (section 4.4).

4.1 Experimental Setup

All experiments were conducted on synthetic data in order to isolate and clearly demonstrate the effect of the proposed operator. Specifically, the synthetic dataset generated at Technische Universität Chemnitz, has been employed. This data provides stereo image pairs with dense ground-truth disparity annotations under controlled rendering conditions. The use of synthetic data guarantees perfect calibration and accurate ground truth, which is essential for isolating and analyzing the geometric contribution of EACS. The primary goal of this evaluation is therefore a controlled proof of concept and a cause–effect analysis—demonstrating that the proposed projection operator behaves as theoretically expected when applied to high-quality flow fields. A comprehensive performance assessment on large-scale public benchmarks is beyond the scope of this initial validation and will be conducted in future work..

Disparity accuracy was quantified using complementary error measures. The Mean Absolute Error (MAE) captures the average deviation between estimated and ground-truth disparity, providing a measure of overall accuracy. The Root Mean Square Error (RMSE) additionally penalizes larger deviations, thereby serving as a robustness indicator against outliers. MAE and RMSE are reported in pixels. To ensure comparability with standard stereo benchmarks, the KITTI D1-all criterion is used as defined by the KITTI protocol, which measures the fraction of pixels whose disparity error exceeds both three pixels and five percent of the ground-truth value. Furthermore, outlier rates beyond fixed thresholds of three and five pixels were computed to provide an additional perspective on robustness. Finally, runtime performance was recorded as the average inference time per frame, including both RAFT flow estimation and the EACS post-processing step, in order to assess the computational efficiency of the proposed pipeline.

4.2 Quantitative Results

The quantitative evaluation demonstrates that the RAFT + EACS pipeline is capable of producing highly accurate disparity estimates from optical flow with sub-pixel precision. On the TU Chemnitz synthetic dataset, the system achieved a Mean Absolute Error (MAE) of 0.3007 pixels and a Root Mean Square Error (RMSE) of 0.9470 pixels. The relatively small difference between these two measures indicates that large errors were rare and that the majority of deviations from ground truth remained small and evenly distributed.

Complementing these findings, the D1-all metric yielded an error rate of only 0.4%, confirming that the fraction of pixels suffering from significant errors was very limited. Threshold-based robustness analysis further revealed that 0.3% of pixels exceeded the three-pixel error margin, while fewer than 0.05% exceeded the five-pixel margin. These values provide strong evidence that the proposed method remains stable even under stringent error definitions. The quantitative performance of RAFT + EACS, together with several baseline comparisons, is summarized in Table 1.

From a computational perspective, the introduction of EACS had negligible impact on runtime performance. The complete pipeline, executed on an NVIDIA RTX 3090 GPU, achieved an average throughput of approximately 10 frames per second at full resolution. The EACS operator itself required less than 0.01 milliseconds per frame, underscoring its suitability for real-time applications where efficiency is critical. Considering the evaluation metrics jointly provides further insight into the behavior of the proposed method. The low MAE indicates high overall accuracy, while the relatively small gap between MAE and RMSE suggests that large disparity errors are rare. This observation is reinforced by the very low D1-all score and threshold-based outlier rates, confirming that most pixels remain well within strict error bounds. Together, these metrics indicate that EACS does not merely improve average accuracy, but also effectively suppresses extreme failure cases by removing geometrically irrelevant flow components.

Table 1. Quantitative comparison of disparity estimation methods on the TU Chemnitz synthetic dataset. RAFT + EACS achieves sub-pixel accuracy with negligible runtime overhead, demonstrating that disparity can be effectively recovered from optical flow without a dedicated stereo network.

Method	MAE (px)	RMSE (px)	D1-all (%)	>3 px (%)	>5 px (%)	FPS
Naïve Channel Selection	0.4121	1.2815	1.2	1.0	0.4	10
No Selection (Flow Magnitude)	0.6894	2.0472	3.8	3.4	1.2	9
RAFT-Stereo (reference)	0.2105	0.8129	0.3	0.2	0.05	5.5
RAFT + EACS (ours)	0.3007	0.9470	0.4	0.3	0.05	10

4.3 Qualitative Results

Beyond numerical evaluation, qualitative inspection provides additional insight into the behavior of RAFT + EACS. Representative examples, illustrated in Fig 4, compare estimated disparity maps with their ground-truth counterparts. The visual results confirm that the proposed approach successfully preserves fine structural details and produces globally consistent disparities across entire scenes. Object boundaries are well preserved, and thin structures are reconstructed with high fidelity.

An especially notable observation is the robustness of the method in texture-poor regions, where traditional disparity estimation often struggles. By discarding the orthogonal component of optical flow and retaining only the epipolar-aligned channel, EACS eliminates spurious variations that might otherwise lead to irregularities or ghosting artifacts. The resulting disparity maps appear smooth and geometrically consistent, further corroborating the quantitative findings.

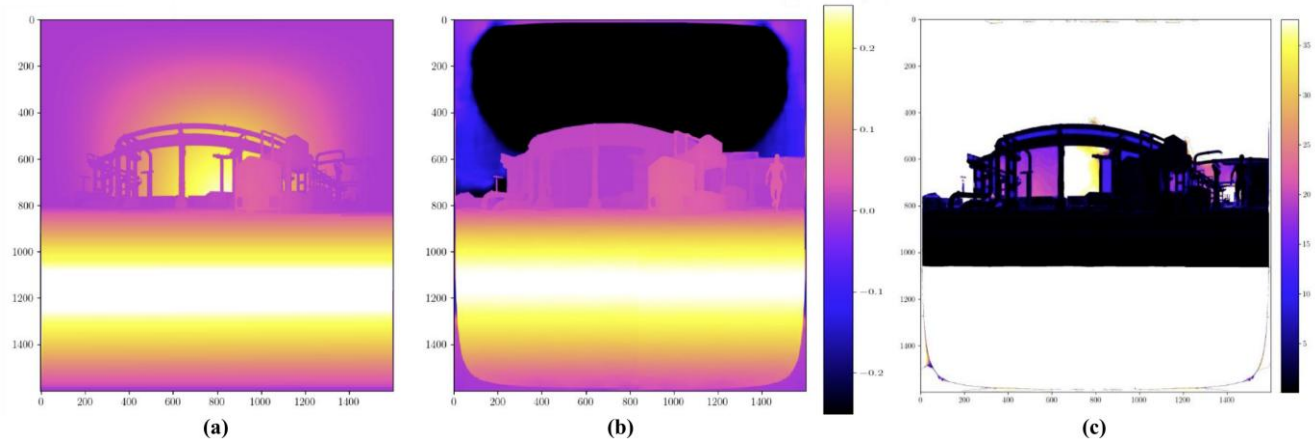

Fig 4: Qualitative results of RAFT + EACS disparity estimation. (a) Ground-truth disparity. (b) Estimated disparity map. The black region in the estimated disparity corresponds to the sky, where disparity is undefined. (c) Pixelwise RMSE error map, where dark colors indicate errors close to zero. Together, the results demonstrate that RAFT + EACS produces disparity maps that are highly consistent with ground truth, with minimal error concentrated only at fine object boundaries.

Representative qualitative results are presented in Fig 4. Subfigure (a) shows the ground-truth disparity, while (b) illustrates the estimated disparity obtained using the RAFT + EACS pipeline. The visual comparison demonstrates that the predicted disparity map closely follows the ground truth, including fine structural details and smooth surfaces. The black region in the estimated disparity corresponds to the sky, where disparity is undefined and can thus be disregarded. Subfigure (c) presents the pixelwise RMSE error map, which is predominantly black, indicating errors close to zero across most of the scene. Only small regions near object boundaries exhibit non-negligible error, confirming that EACS effectively extracts the geometrically meaningful disparity signal.

4.4 Discussion

Taken together, the experimental results validate the core claim of this work, i.e. accurate stereo disparity can indeed be obtained directly from optical flow by removing the non-epipolar channel. Despite its conceptual simplicity, the EACS operator consistently produced results that were both numerically precise and visually convincing. The strong performance can be attributed to two complementary factors. First, RAFT provides dense and accurate flow fields in which the disparity signal is already implicitly encoded. Second, the explicit enforcement of epipolar geometry through channel selection ensures that this signal is cleanly extracted without interference from irrelevant components.

A closer examination of Table 1 reveals a clear trade-off between architectural specialization and computational efficiency. While RAFT-Stereo achieves slightly lower error rates, it does so at the cost of a dedicated stereo architecture and reduced inference speed. In contrast, RAFT + EACS achieves competitive accuracy using a generic optical flow backbone and a parameter-free projection step, effectively closing much

of the performance gap while nearly doubling runtime throughput. The performance difference can be attributed to the fact that RAFT-Stereo explicitly optimizes a disparity-aligned cost volume during training, whereas RAFT + EACS relies solely on the quality of the underlying optical flow. Importantly, the results demonstrate that a large portion of the disparity signal is already present in standard optical flow representations and can be recovered through geometry-aware post-processing alone.

These findings carry important implications for the design of future disparity estimation systems. They demonstrate that dedicated stereo networks are not strictly necessary when high-quality optical flow is available, since RAFT + EACS achieves competitive results without retraining or architectural modifications. At the same time, the results highlight the limitations of the approach: Because the method builds directly on RAFT's optical flow, any flow inaccuracies—such as those introduced by occlusions, motion discontinuities, or reflective surfaces—also appear in the resulting disparity map. This behavior is not unique to our approach; most stereo methods experience similar challenges in these regions due to the inherent ambiguity of correspondence estimation. . Furthermore, the experiments were conducted exclusively on synthetic data under perfect calibration, and extending validation to real-world datasets remains an important next step. Finally, while EACS is readily applicable to rectified stereo pairs with horizontal or vertical baselines, its generalization to arbitrary epipolar orientations will require further extensions, such as locally adaptive projections.

5. CONCLUSION AND FUTURE WORK

This study has introduced Epipolar-Aligned Channel Selection (EACS), a lightweight post-processing operator designed to extract disparity directly from optical flow by enforcing epipolar constraints. In contrast to specialized stereo networks, the proposed method adds no trainable parameters and introduces negligible computational overhead, yet it successfully converts dense optical flow fields into disparity maps that closely align with ground truth. By coupling EACS with RAFT, this paper demonstrated that accurate disparity can be achieved without the need for retraining or architectural modifications. Experimental results on the TU Chemnitz synthetic dataset confirm the claim that the RAFT + EACS pipeline consistently delivered sub-pixel disparity accuracy, with a Mean Absolute Error of 0.3007 pixels, a Root Mean Square Error of 0.9470 pixels, and exceptionally low error rates under strict benchmarks (D1-all = 0.4%). The qualitative evaluation further highlighted the method's ability to preserve fine structural details, avoid spurious noise, and remain robust in low-texture regions where disparity estimation is generally challenging.

The broader implication of these findings is that geometry-aware post-processing can serve as a powerful alternative to network retraining in stereo disparity estimation. The proposed approach demonstrates that when a high-quality optical flow estimator is available, the disparity signal is already embedded within one channel of the flow representation. Rather than explicitly encoding geometric knowledge into the network, the method leverages prior geometric understanding to guide the design of the post-processing operator—specifically, by selecting the disparity-aligned flow component based on epipolar geometry. This perspective bridges classical vision principles with modern deep architectures without modifying or retraining the underlying network.

Nevertheless, the study also uncovers several limitations that warrant further investigation. First, because the approach is entirely dependent on the accuracy of RAFT's optical flow, errors introduced by occlusions, motion discontinuities, or reflective surfaces inevitably propagate into the disparity output. Second, the current evaluation has been restricted to synthetic data under ideal calibration, and extending the analysis to real-world benchmarks such as KITTI or Middlebury is a necessary step to assess robustness in practical conditions. Third, the present formulation of EACS assumes a globally fixed baseline orientation (horizontal or vertical). While this assumption holds for most rectified stereo setups, generalizing the operator to arbitrary or spatially varying baselines remains an open challenge. In practice, when the baseline is horizontal, only the horizontal component of the flow (f_x) is preserved, whereas in vertical stereo rigs the selection is reversed to retain f_y. For more general stereo geometries (e.g., wide-baseline or oblique rigs), EACS could be extended to select a linear combination of the horizontal and vertical flow components (f_x, f_y) (i.e. $f_{EACS}(i,j) = e^T . f(i,j)$= $e_x . f_x + e_y . f_y$) , projected along the baseline vector e (optionally local, $e(i,j)$, for spatially varying baselines), thereby maintaining alignment with the underlying epipolar geometry. Addressing this may require locally adaptive extensions of EACS, potentially supported by lightweight trainable modules that can dynamically align flow with the baseline direction.

Looking ahead, several promising research directions emerge. One avenue involves integrating RAFT + EACS into broader multi-task frameworks for scene flow, depth–motion estimation, or autonomous navigation pipelines. Because EACS is both differentiable and parameter-free, it could be seamlessly combined with existing architectures, providing an efficient preprocessing step that reduces redundancy in downstream learning. Another direction lies in exploring fine-tuning strategies, such as knowledge distillation or domain adaptation, to further improve flow-to-disparity conversion in challenging real-world scenarios. Finally, extending the evaluation to dynamic environments with moving objects, variable lighting conditions, and sensor imperfections would provide valuable insight into the method's applicability under operational constraints.

An important direction for future work is the extension of the experimental evaluation to established public benchmarks such as KITTI [33], Middlebury [34], and SceneFlow [13], which introduce real-world challenges including imperfect calibration, occlusions, and varying lighting conditions. In addition, evaluating the method under different stereo configurations, baselines, and wide-FOV scenarios would further assess the robustness and generality of EACS. Such evaluations would complement the current controlled study and provide a comprehensive picture of the method's performance across diverse real-world scenarios.

In conclusion, this work establishes that accurate stereo disparity can be recovered directly from optical flow through a simple, geometry-aware channel selection step. The results highlight the strength of combining state-of-the-art flow estimation with explicit epipolar priors, showing that principled post-processing can substitute for specialized networks in many contexts. By bridging the gap between optical flow and disparity with a minimal yet effective operator, the proposed framework opens new opportunities for efficient, real-time 3D vision in robotics, autonomous systems, and immersive media applications.

6. REFERENCES

[1] Vedula, S., Baker, S., Rander, P., Collins, R., and Kanade, T. 1999. Three-dimensional scene flow. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). 722–729.

[2] Horn, B. K. and Schunck, B. G. 1981. Determining optical flow. Artificial Intelligence. 17, 1–3, 185–203.

[3] Zach, C., Pock, T., and Bischof, H. 2007. A duality based approach for realtime TV-L1 optical flow. In Proceedings of the DAGM Conference on Pattern Recognition. 214–223.

[4] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., and Brox, T. 2015. FlowNet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2758–2766.

[5] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., and Bry, A. 2017. End-to-end learning of geometry and context for deep stereo regression. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). 66–75.

[6] Chang, J.-R. and Chen, Y.-S. 2018. Pyramid stereo matching network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 5410–5418.

[7] Sun, D., Yang, X., Liu, M.-Y., and Kautz, J. 2018. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 8934–8943.

[8] Poggi, M., Tosi, F., Batsos, K., Mordohai, P., and Mattoccia, S. 2020. On the synergies between machine learning and binocular stereo for depth estimation from images: A survey. arXiv preprint arXiv:2004.08566.

[9] Teed, Z. and Deng, J. 2020. RAFT: Recurrent all-pairs field transforms for optical flow. In Proceedings of the European Conference on Computer Vision (ECCV). 402–419.

[10] Lipson, L., Teed, Z., and Deng, J. 2021. RAFT-Stereo: Multilevel recurrent field transforms for stereo matching. In Proceedings of the IEEE International Conference on 3D Vision (3DV). 218–227.

[11] Liu, X., Zhang, T., and Liu, M. 2024. Joint estimation of pose, depth, and optical flow with a competition–cooperation transformer network. Neural Networks 171, 263–275.

[12] Guo, X., Zhao, H., Shao, S., Li, X., and Zhang, B. 2024. F²Depth: Self-supervised indoor monocular depth estimation via optical flow consistency and feature map synthesis. Neural Networks 175, 106–118.

[13] Gui, M., Schusterbauer, J., Prestel, U., Ma, P., Kotovenko, D., Grebenkova, O., Baumann, S. A., Hu, V. T., and Ommer, B. 2024. DepthFM: Fast generative monocular depth estimation with flow matching. arXiv preprint arXiv:2403.13788.

[14] Birchfield, S. and Tomasi, C. 1998. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 4, 401–406.

[15] Zbontar, J. and LeCun, Y. 2016. Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research 17, 1, 2287–2318.

[16] Mayer, N., Ilg, E., Häusser, P., Fischer, P., Cremers, D., Dosovitskiy, A., and Brox, T. 2016. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 4040–4048.

[17] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., and Brox, T. 2015. Flownet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2758–2766.

[18] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., and Brox, T. 2017. Flownet 2.0: Evolution of optical flow estimation with deep networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2462–2470.

[19] Zhang, F., Prisacariu, V., Yang, R., and Torr, P. H. 2019. GA-Net: Guided aggregation net for end-to-end stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 185–194.

[20] Guo, X., Yang, K., Yang, W., Wang, X., and Li, H. 2019. Group-wise correlation stereo network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 3273–3282.

[21] Hur, J. and Roth, S. 2019. Iterative residual refinement for joint optical flow and occlusion estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 5754–5763.

[22] Teed, Z. and Deng, J. 2021. RAFT-3D: Scene flow using rigid-motion embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 8375–8384.

[23] Mehl, L., Jahedi, A., Schmalfuss, J., and Bruhn, A. 2023. M-fuse: Multi-frame fusion for scene flow estimation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2020–2029.

[24] Jeong, J., Lin, J. M., Porikli, F., and Kwak, N. 2022. Imposing consistency for optical flow estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 3181–3191.

[25] Huguet, F. and Devernay, F. 2007. A variational method for scene flow estimation from stereo sequences. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). 1–7.

[26] Wedel, A., Brox, T., Vaudrey, T., Rabe, C., Franke, U., and Cremers, D. 2011. Stereoscopic scene flow computation for 3D motion understanding. International Journal of Computer Vision 95, 1, 29–51.

[27] Poggi, M., Tosi, F., Batsos, K., Mordohai, P., and Mattoccia, S. 2021. On the synergies between machine learning and binocular stereo for depth estimation from images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[28] Zhai, M., Xiang, X., Lv, N., and Kong, X. 2021. Optical flow and scene flow estimation: A survey. Pattern Recognition 114, 107861.

[29] Ilg, E., Saikia, T., Keuper, M., and Brox, T. 2018. Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation. In Proceedings of the European Conference on Computer Vision (ECCV). 614–630.

[30] Jiang, H., Sun, D., Jampani, V., Lv, Z., Learned-Miller, E., and Kautz, J. 2019. SENSE: A shared encoder network for scene-flow estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 3195–3204.

[31] Jaegle, A., Borgeaud, S., Alayrac, J.-B., Doersch, C., Ionescu, C., Ding, D., Koppula, S., Zoran, D., Brock, A., Shelhamer, E., et al. 2021. Perceiver IO: A general architecture for structured inputs and outputs. arXiv preprint arXiv:2107.14795.

[32] Li, J., Wang, P., Xiong, P., Cai, T., Yan, Z., Yang, L., Liu, J., Fan, H., and Liu, S. 2022. Practical stereo matching via cascaded recurrent network with adaptive correlation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 16263–16272.

[33] Geiger, A., Lenz, P., and Urtasun, R. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3354–3361.

[34] Scharstein, D. and Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47, 1–3, 7–42.