# Cross-based Matching Constrained by the Classes of Pixels

Yang, Y.C. [1*] and Jaw, J.J. [2]

[1]PhD student, Department of Civil Engineering, National Taiwan University, Taiwan

[2]Associate Professor, Department of Civil Engineering, National Taiwan University, Taiwan

*ivy.ycyang@gmail.com

***Abstract:*** *In the field of computer vision, accurate depth estimation is crucial for various applications such as 3D reconstruction, object recognition, and autonomous navigation. This paper presents an optimized approach to stereo matching that integrates Cross-Based Matching and image segmentation. In this research, a classified image is generated by pixel classification to enhance the precision of disparity maps. This method leverages Semi-Global Matching (SGM) for its robustness and reliability while introducing a unique constraint based on pixel classification. This constraint incorporates image segmentation to inform the Cross-based matching process, setting this approach apart from traditional SGM and Cross-based matching methods. Classifying pixels into distinct categories and using these classifications to restrict the matching area significantly reduces ambiguities and improves consistency in disparity estimation. The experiment used different numbers of objects in image segmentation to perform several tests. The results of the proposed method versus Cross-based matching were evaluated using five metrics: difference map, error evaluation, error rate, optimized disparity, and error distribution map. Experimental results demonstrate that the proposed method indeed reduces the error in Cross-based matching, particularly distributed around the borders of the different objects. However, some new errors arise because of insufficient quality of classification. This integration of Cross-based with image classification constraints provides an ideal path in stereo matching techniques, paving the way for more accurate and reliable depth estimation in various computer vision applications. Yet, how to best employ segmentation information and quality for improving stereo matching remains not only interesting but also challenging.*

*Keywords: SGM, cross-based matching, image classification, classified image, disparity map*

## Introduction

In computer vision, accurate depth estimation plays a crucial role in various applications such as 3D reconstruction, object recognition, and autonomous navigation. Among the techniques used for depth estimation, stereo matching is one of the most fundamental and widely researched approaches. Stereo matching involves generating a disparity map by comparing pixel correspondences between two images taken from slightly different viewpoints. This disparity map is then used to infer the depth of the information in the scene. However, traditional stereo-matching methods often struggle with regions of low texture, occlusions, and repetitive patterns, leading to inaccurate depth estimates in such areas.

Cross-based matching is a technique that has shown promise in addressing some of these challenges by employing local cross-shaped support regions to improve pixel correspondence.

Despite its effectiveness, Cross-based matching can still face difficulties when encountering complex image regions, particularly at object boundaries where matching ambiguities are more pronounced. To improve the precision of disparity maps and mitigate these limitations, recent research has explored the integration of additional information, such as pixel classification and image segmentation, into the stereo-matching process.

This paper proposes an optimized stereo-matching approach that combines Cross-Based Matching with image segmentation, leveraging pixel classification to inform and constrain the matching process. The method also integrates Semi-Global Matching (SGM) for its robustness in handling challenging regions while maintaining computational efficiency. By classifying pixels into distinct categories, such as edges, textures, and flat regions, and introducing these classifications as constraints in the Cross-based matching process, we aim to reduce ambiguities and improve consistency in disparity estimation, particularly around object borders. The core innovation of this research lies in the use of pixel classification to guide Cross-based Matching. This novel constraint-based approach restricts the matching search space for each pixel class, thereby enhancing the accuracy of disparity estimation in complex image regions. For example, edge pixels are constrained to ensure strong spatial consistency along object boundaries, while texture and flat region pixels are matched with constraints that reduce errors typically encountered in traditional methods. The integration of segmentation-based constraints into stereo matching sets our approach apart from traditional Cross-based matching and SGM methods.

While the proposed method shows promise in improving disparity map accuracy, particularly at object boundaries, challenges remain. One notable challenge is the quality of the image segmentation itself. Poor segmentation can introduce new errors, especially in regions where pixels are misclassified. Despite these challenges, this research provides a valuable step forward in stereo matching techniques by integrating Cross-Based Matching with image segmentation constraints, offering a more reliable and accurate solution for depth estimation in a wide range of computer vision applications.

Furthermore, the application of this method to photogrammetric point clouds is explored in this study. Photogrammetric point clouds, generated from image-based techniques, often suffer from several challenges: they can be unstructured, lack important attribute information, and involve massive datasets that are computationally demanding to manage. By enriching point clouds with attribute data derived from stereo matching and pixel classification, this research seeks to address these limitations, transforming point cloud data into more structured and meaningful point cloud information.

Thus, this paper not only aims to enhance stereo matching by introducing constraints based on pixel classification but also contributes to the field of photogrammetry by improving the information content of point clouds. Through a series of experiments, the effectiveness of the proposed method is evaluated using several metrics, including pixel accuracy, difference maps, and error distribution analysis, demonstrating its potential to provide more accurate and reliable depth estimation.

## Literature Review

### a. Stereo Matching

Stereo matching is a crucial technique in computer vision used to estimate depth from two or more images taken from slightly different viewpoints. The key objective is to match corresponding pixels between the stereo images to calculate disparity, which provides depth information.

Early methods, such as Sum of Absolute Differences (SAD) (Hamzah, R. A., Abd Rahim, R., & Noh, Z. M., 2010) and normalized cross-correlation (NCC) (Yoo & Han, 2009), were pixel-based or block-based techniques. These methods, while simple, often struggled with occlusions, textureless regions, and illumination differences, particularly around object boundaries.

Global stereo-matching methods, like those using Graph Cuts (Hong & Chen, 2004) or Belief Propagation (Yang et al., 2006), improved accuracy by considering relationships across the entire image, though at a high computational cost. SGM, introduced by Hirschmüller (2005), struck a balance by offering better accuracy than local methods and more efficiency than fully global methods. SGM became widely adopted for its effectiveness in practical applications.

### b. Semi-Global Matching

SGM is a widely used algorithm in computer vision and image processing for stereo matching, particularly in depth estimation from stereo images (Hirschmuller, 2005). It aims to compute corresponding points in two images taken from different viewpoints, which helps reconstruct a 3D scene. The basic idea is to find correspondences between pixels in stereo images by minimizing a cost function. This cost function evaluates the similarity between pixels or image patches in the left and right images. SGM differs from traditional local stereo matching algorithms by considering the matching cost of individual pixels and incorporating information from neighboring pixels and multiple scanline directions. In SGM, the cost aggregation process involves computing a cost volume that contains matching costs for all possible disparities (horizontal offsets) for each pixel in the based image. This cost volume is then efficiently minimized using dynamic programming techniques. By considering costs along multiple paths, including horizontal, vertical, and diagonal directions, SGM produces more accurate and robust disparity maps than purely local methods.

**c. Cross-based matching**

SGM is highly sensitive to the penalty parameter, and the results before and after adding the penalty parameter are significantly different at the edge of the object. Cross-based matching attempts to address some of these issues by using local cross-shaped support regions around each pixel. This algorithm considers each pixel when determining the summed range of matching cost values to establish shape-adaptive support regions of different sizes based on the color similarity and connectivity between its neighboring pixels (Zhang et al., 2009). The cross shape allows cross-based matching to maintain local consistency in matching, particularly in areas of the image where traditional methods would introduce noise. Cross-based matching can provide a set of initial disparity maps before performing SGM. Then use it to generate the confidence map which would be used to automated decide the penalty parameters for vary dataset. So that the subsequent SGM can smooth the image edges. It can reduce the sensitivity of the penalty parameter and correct the non-smoothness and parallax discontinuity (Ting & Jaw, 2017).

**d. Point Clouds Classification**

Currently, point cloud classification is typically performed after the point cloud is generated. Yang et al. (2022) proposed two deep learning models, Geometric relation-based convolution (GRC) and relational attention interpolation (RAI), for point cloud classification and segmentation. Similarly, Pessoa et al. (2019) introduced a pixel-supervised classification method based on decision trees, which uses the mean and standard deviation of class attributes extracted from training samples to define the decision tree, applying it to data validation and accuracy assessment. This research trend means that large amounts of unprocessed 3D data must be handled simultaneously during this process. This can lead to challenges in data processing. The main limitation of this approach is that the unprocessed 3D data may contain significant noise, missing values, or unnecessary details, which can negatively impact the performance of the classification models. If effective preprocessing of the optical images used to generate the point cloud can be performed before classification, it would help improve the accuracy and efficiency of the classification models, ensuring that they focus only on the data features most relevant to the task.

**e. Main Goal**

One solution to the limitations of cross-based matching is to incorporate additional information through pixel classification and image segmentation. By classifying pixels into different types, such as edges, textures, or flat regions, this classification can inform stereo matching. Our classification-based approach ensures that pixels are matched according to their specific characteristics, thereby reducing matching ambiguities and improving overall consistency in disparity estimation.

This research proposes an innovative combination of cross-based matching with image classification to refine stereo matching. Classifying pixels into distinct categories enables the application of class-specific constraints during the matching process, providing improved precision in disparity maps, particularly at object edges.

**Methodology**

The proposed method, referred to as "Class-aided" throughout this paper, integrates Cross-based matching with pixel classification to provide an optimized initialized disparity map for the subsequent SGM. Class-aided matching leverages semantic class information and integrates it into cost aggregation in Cross-based matching to guide the image-matching process in a more reliable way. Considering the underlying object classes in the scene, this approach enhances the matching reliability, particularly in areas where traditional methods struggle. Specifically, the methodology integrates manual image classification images, referred to as classified images, into Cross-based matching and generates an initialized disparity to subsequent processes. The classified image extracts high-level semantic features, which are then used to guide the matching cost computation and initialized disparity map refinement. Class-aided matching involves classifying pixels based on image classification to constrain the cost aggregation process, enhancing depth estimation and disparity map accuracy. The steps in the methodology are outlined in Figure 3:

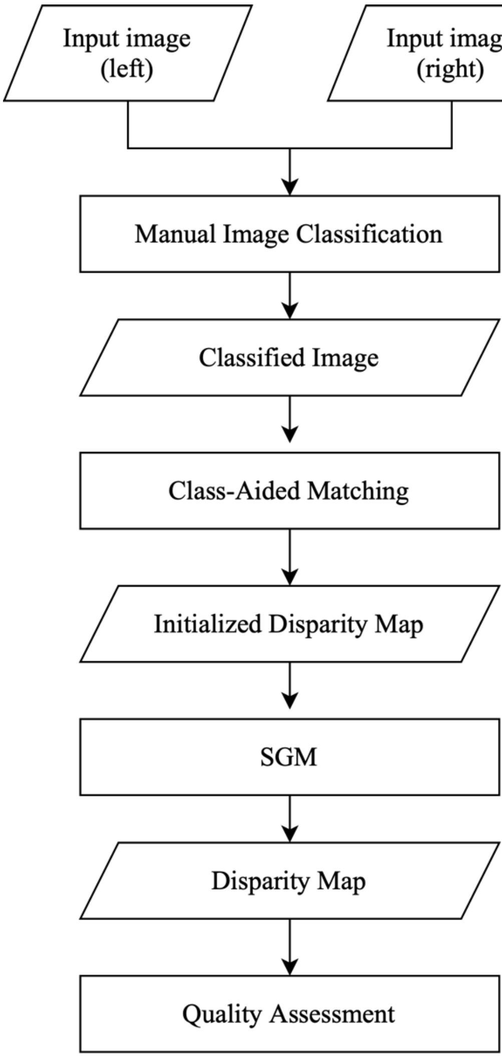



Figure 1: Workflow

**a. Dataset**

The image pair used in the experiment comes from the Middlebury dataset (Scharstein & Szeliski, 2003). It consists of high-resolution stereo sequences with complex geometry and provides pixel-accurate ground truth of disparity and occlusion area that could be used to evaluate the quality of the result. The image pair is shown in Figure 8.

(Cones)

(Teddy)

Figure 2: Experiment image pair

**b. Experiment configuration**

The experiments were conducted using multiple configurations, including varying the number of objects in the image classification and adjusting the classification quality. These parameters allowed us to analyze the sensitivity of the method to different classification accuracies and segmentation complexities. Detailed description is shown in Table 1. By following this experimental configuration, Class-aided matching was rigorously evaluated to demonstrate its potential to enhance disparity map accuracy while addressing the limitations of traditional stereo-matching techniques.

Table 1: Experiment configuration

| | |
|---|---|
| **Cross-based** | Used as the reference method to assess the performance impact of different Class-aided matching configurations. |
| **Class-aided** | Involves classifying all objects in the scene and assigning distinct labels to each. |
| **Partial Class-aided** | Consolidates labels in areas where adding labels caused frequent new errors, aiming to reduce misclassifications. |
| **Ground truth Class-aided** | Refines the classified image based on ground truth disparity data. This method prioritizes the left image, checks for discrepancies in the right image and its labels, and makes manual adjustments accordingly. |
| **Partial Ground truth Class-aided** | The same treatment method as Ground truth Class-aided but merging labels for certain regions, such as the background. Instead of applying the ground truth data across the entire image, it focuses on key areas where errors are common, like object boundaries. |

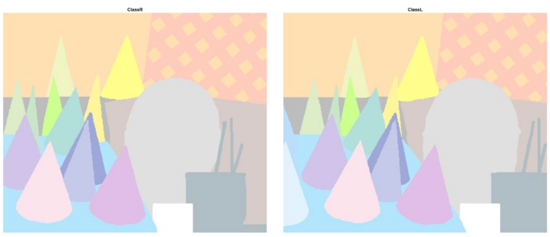
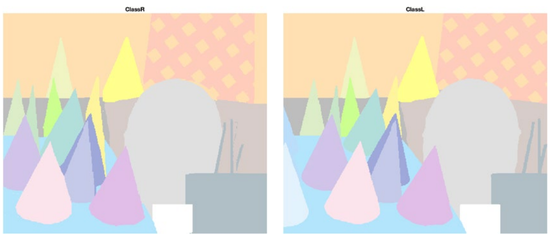
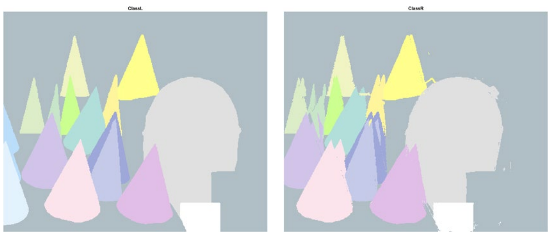
**c. Classified image**

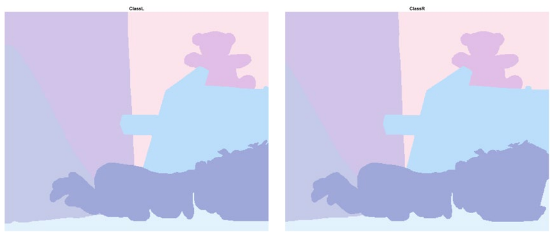
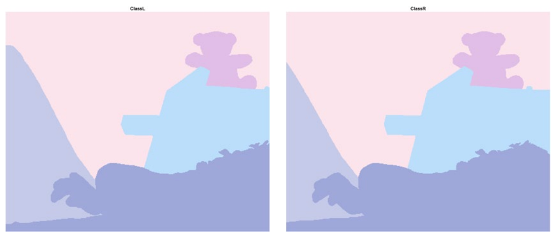
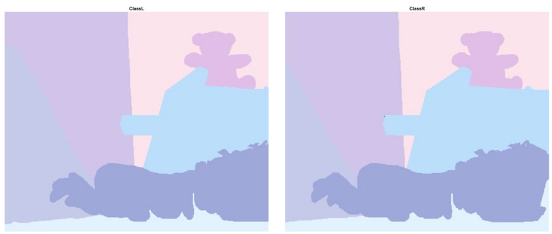
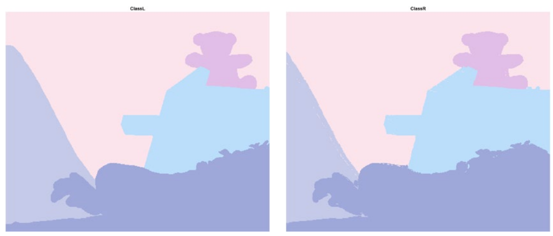
An image classification process is employed before conducting stereo matching. Each pixel on the classified image is assigned a class label that corresponds to its local image characteristics. These class labels will be used in the subsequent stereo-matching process to enforce class-specific potential matching areas. Limit the cost aggregation area and compute the pixel cost only when the labels of the potential correspondence points are identical.

Figure 5 is the initial result after using a pre-trained model, Segment Anything Model (SAM) (Kirillov et al., 2023), which is only used for image segmentation and has yet to add semantic information. It can be observed that the result is acceptable in the overall area, but there are still some correspondence places that have different segmentation results(red circles in Figure 3). If these classified images are used in the sequence experiment process, then this issue is an essential task to solve. Thus, to evaluate the effect of the proposed algorithms, a manually image classification was applied provisionally to the research, the classified image used in the experiments are shown in Table 2.

Figure 3: Initial classified image

Table 2: Manual classified image

| **Cones** | |
|---|---|
| Class-aided | Partial Class-aided |
|  |  |
| Ground truth Class-aided | Partial Ground truth Class-aided |
|  |  |

| **Teddy** | |
|---|---|
| Class-aided | Partial Class-aided |
|  |  |
| Ground truth Class-aided | Partial Ground truth Class-aided |
|  |  |

### d. Class-Aided Matching

In Class-aided methods, classified image is utilized as a constraint to restrict the range of stereo matching. The classification information, which provides semantic labels for each pixel, serves as a guide in the initialized disparity map generation by introducing pixel-wise constraints. Specifically, it ensures that only pixels belonging to the same semantic class are matched while discarding matches between pixels with different labels. This approach helps refine the matching process, particularly in regions where Cross-based matching struggles, such as around object boundaries.

The following pixel cost calculation functions illustrate how classification constraints are incorporated into the algorithm as shown below equations:

$$C_{AD}(p,d) = \begin{cases} \frac{1}{3}\sum_i \left| I_i^{left}(p) - I_i^{right}(p,d) \right|, & Class_L(p) = Class_R(p,d) \\ \infty, & Class_L(p) \neq Class_R(p,d) \end{cases} \quad (1)$$

$$C_{Census}(p,d) = \begin{cases} \frac{1}{3}\sum_i H\left(Census_L(p) - Census_R(p,d)\right), & Class_L(p) = Class_R(p,d) \\ \infty, & Class_L(p) \neq Class_R(p,d) \end{cases} \quad (2)$$

$$Census(p) = Bitstring_{q \in Np}\left(I(q) \geq I(p)\right)$$

$$\forall i \in \left\{ \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, RGB - based \right\}$$

where

$C_{AD}(p,d)$: AD matching cost.

$I_i^{left}(p)$: The gradient value of the pixel in the left image.

$I_i^{right}(p,d)$: The gradient value of the pixel in the left image corresponding to the right image, with disparity $d$.

$Class_L(p)$: The label of the pixel in the left image.

$Class_R(p,d)$: The label of the pixel in the left image corresponding to the right image, with disparity $d$.

$C_{Census}(p,d)$: Census matching cost.

$H$: Hamming distance.

$I(p)$: Color gradient values of adjacent pixels within the mask.

$I(q)$: Color gradient value of the center pixel in the mask.

$\left\{\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, RGB - based\right\}$: The image gradient values of the three RGB bands in the x and y directions respectively.

This method integrates the classified image information, leveraging it to enforce more accurate matches by only considering pixels of the same class. This significantly improves the algorithm's robustness in challenging regions like object boundaries, reducing errors and improving depth estimation performance. The labeling pixels are treated with stronger spatial constraints, reducing the chances of mismatches along object borders. This combined approach demonstrates the potential of using classification constraints to improve the quality of disparity maps, ultimately providing a more robust solution for depth estimation in stereo-matching tasks.

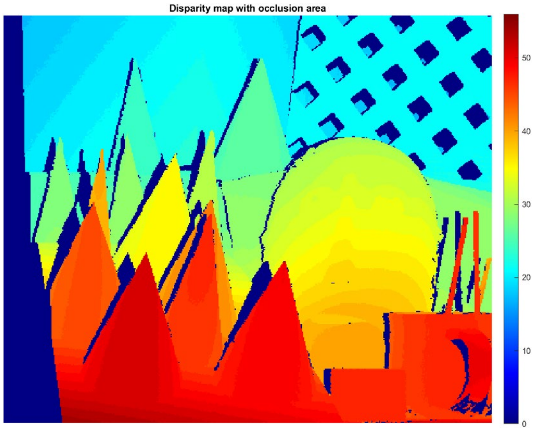
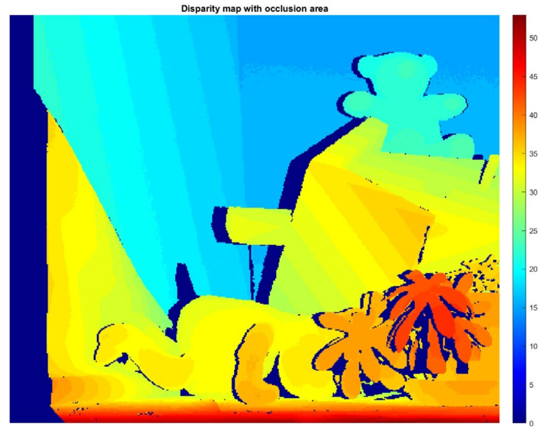
**e. Quality Assessment**

Finally, a quality assessment step is performed to evaluate the accuracy of the generated disparity maps. Several evaluation metrics were employed to assess the performance of the proposed Class-aided matching method comprehensively. Each of these metrics plays a crucial role in quantifying the accuracy of disparity map generation and identifying potential areas for improvement. The following sections detail the key evaluation metrics used in the experiment:

- **Error rate**

Pixel accuracy refers to the percentage of error-matched pixels in the generated disparity map compared to the ground truth of the non-occlusion area (Table 3). It measures the overall performance of the stereo-matching algorithm. Equation 3 is calculated as the ratio of the number of pixels with wrong estimated disparity values to the total number of pixels in the non-occlusion area.

Table 3: Ground truth data of non-occlusion

| **Cones** | **Teddy** |
|---|---|
|  |  |

$$B_{Non-occ} = \frac{1}{N}\left(\sum_{p \in Non-occ} |d(p) - d_{Non-occ}(p)| > \delta\right) \quad (3)$$

where:

$B_{Non-occ}$: The error rate of the non-occlusion area.

$d(p)$: Ground truth of the non-occlusion area.

$d_{Non-occ}(p)$: The disparity map of the non-occlusion area.

$\delta$: Threshold, usually to be 1.

A lower error rate indicates that the stereo-matching algorithm performs well in aligning the depth values, contributing to more reliable disparity maps and reflecting the overall effectiveness of the proposed algorithm in depth estimation.

- **Difference map**

A difference map is a visual representation of the differences between the disparity map and the ground truth, which is calculated by Equation 4. This map highlights the pixel-by-pixel differences, making it easy to identify regions where the algorithm performed poorly. By inspecting the difference map, we can pinpoint specific problem areas, such as occluded regions, object boundaries, and regions with repetitive patterns, which are commonly difficult for stereo-matching algorithms.

$$Diff_{Non-occ} = |d(p) - d_{Non-occ}(p)| \quad (4)$$

- **Optimized disparity**

The optimized disparity is a representation of the improvement achieved by using the Class-aided matching approach compared to the Cross-based method. It is calculated by subtracting the difference map of Class-aided matching from the difference map of Cross-based matching, as shown in Equation 5:

$$O_{Non-occ} = |Diff_{Cross}| - |Diff_{Class}| \tag{5}$$

where

$Diff_{Cross}$: The difference map of Cross-based matching.

$Diff_{Class}$: The difference map of Class-aided matching.

This process highlights the areas where Class-aided matching has reduced the disparity errors relative to the Cross-based method, specifically within the non-occlusion regions. If the value of the optimized disparity is positive, it indicates that Class-aided matching has improved the accuracy in that region. Conversely, if the value is negative or zero, it suggests that there was no improvement or that Class-aided matching performed similarly or worse. This visualization helps to clearly assess the effectiveness of the optimization in refining the disparity map.

- **Optimized area**

Optimized area plays a critical role in determining the effectiveness of the Class-aided matching methods by quantifying the overall disparity errors. The formula can be expressed Equation 6:

$$A_{Non-occ} = \left( \sum_{p \in Non-occ} (|d(p) - d_{Non-occ}(p)| > \delta) * n_{|d(p)-d_{Non-occ}(p)|>\delta} \right) \tag{6}$$

where:

$A_{Non-occ}$: The optimized area of the non-occlusion area.

$n$: the number of pixels where errors occurred

It calculates the optimized area by considering both the number of mismatched pixels in the non-occlusion area and the magnitude of the disparity error for each pixel. It provides a comprehensive measure of the overall error in the disparity map, capturing both the frequency of pixel mismatches and the severity of those errors. Thus, it reflects how many pixels were mismatched and how much the disparity values deviate from the ground truth, giving a deeper insight into the algorithm's performance.

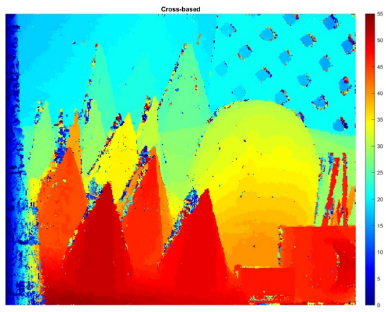
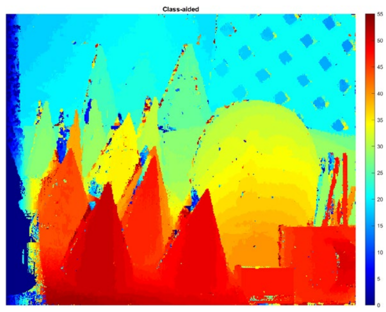
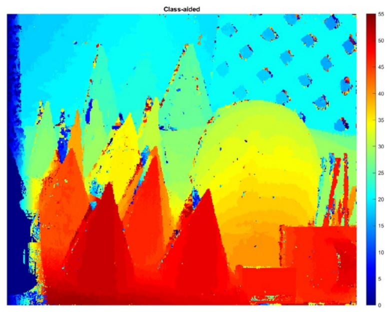
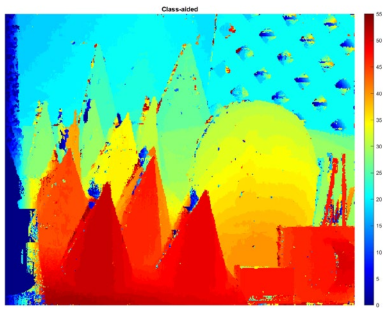
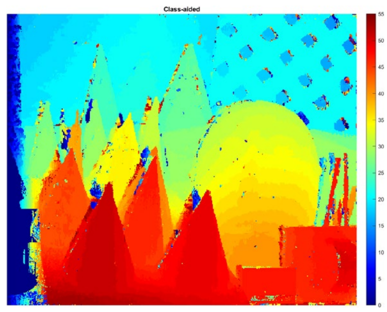
- **Error Distribution Map**

The error distribution map is a spatial representation of the disparity errors within the image. Unlike the difference map, it compares the error pixel location between the reference group (Cross-based matching) and the other test groups. By visually mapping the distribution of errors across the image, researchers can observe how different algorithms perform in specific regions, such as edges, textures, or low-texture areas. The map indicates where errors occur and how their frequency and distribution change when applying different methods. This visual comparison helps identify patterns of disparity errors and assess the effectiveness of various stereo-matching methods, especially in challenging regions of the image.

**Results and Discussion**

In the results section, we will analyze both the initial disparity image and the final disparity map. The initial disparity image is produced after applying Class-aided matching, while the final disparity map is generated after applying SGM, which uses the initial disparity to determine the penalty parameters. Tables 4 and 5 present the initial disparity image and the final disparity map, respectively, providing both visual and quantitative insights into the depth estimation performance of various image-matching methods.

Table 4: Initialized disparity map of different methods

| **Cones** | | |
|---|---|---|
| Cross-based | Class-aided | Partial Class-aided |
|  |  |  |
| Ground truth Class-aided | Partial Ground truth Class-aided | |
|  |  | |

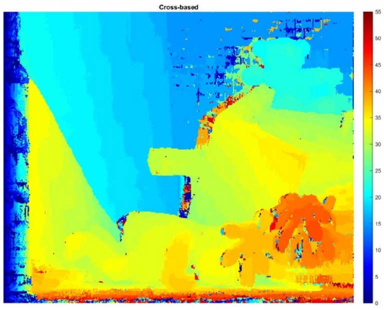
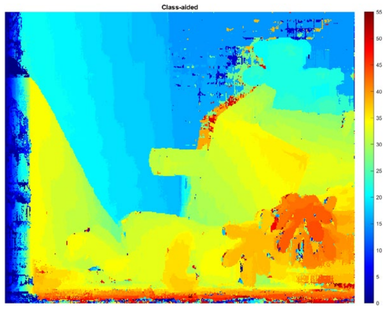
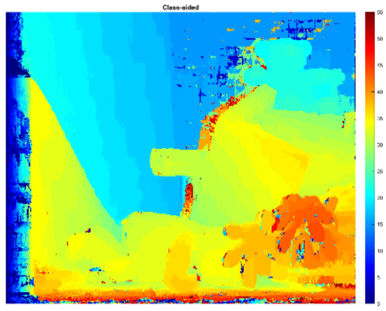
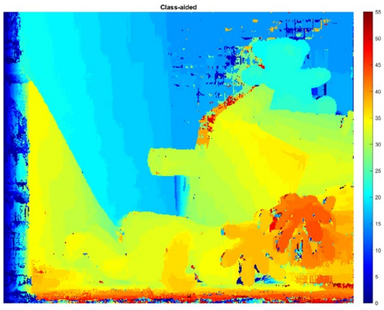
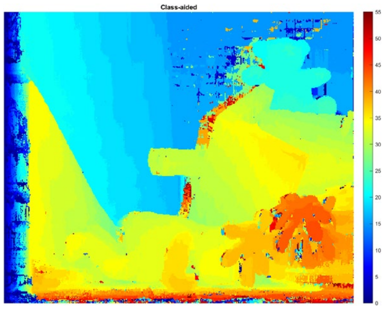
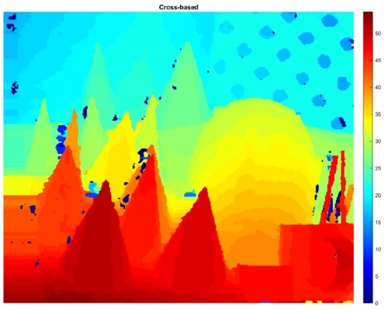
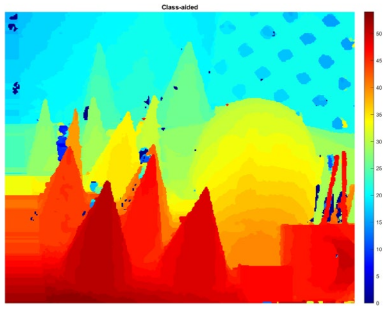
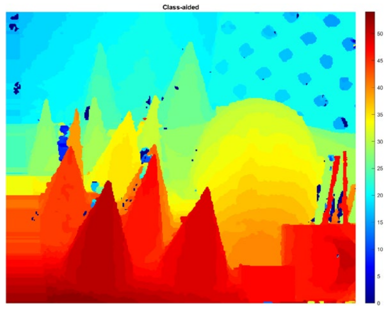
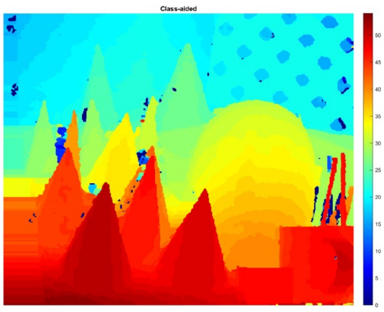
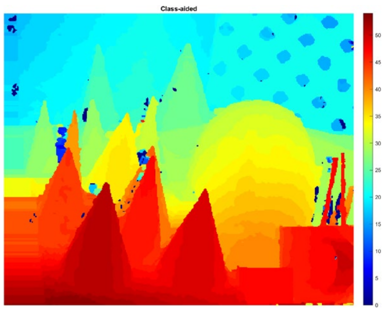
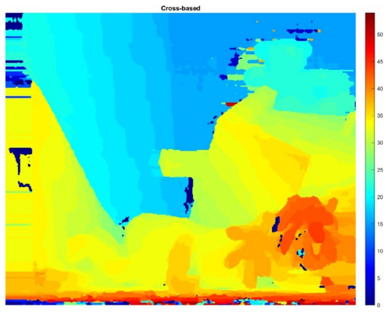
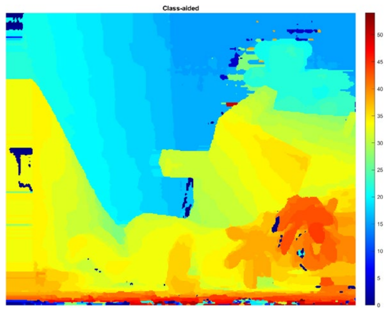
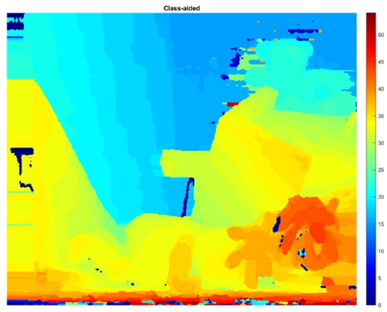
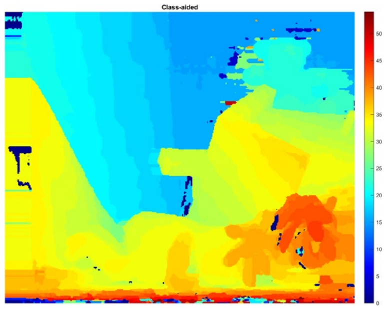
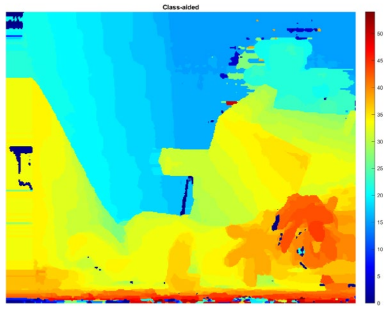
<table>
<tr><th colspan="6">Teddy</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3"></td><td colspan="3"></td></tr>
</table>

Table 5: Final disparity map of different methods

<table>
<tr><th colspan="6">Cones</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3"></td><td colspan="3"></td></tr>
</table>

<table>
<tr><th colspan="6">Teddy</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3"></td><td colspan="3"></td></tr>
</table>

The difference maps of the initial disparity image and the final disparity map in Tables 6 and 7 show that most errors are concentrated around the edges of objects with different labels. This is expected, as object boundaries typically pose challenges for matching algorithms due to variations in texture, lighting, and occlusion. In addition to edge errors in the Teddy dataset, there are noticeable planar errors in the background, where the high texture repeatability or homogeneity in the original image complicates matching. However, aside from these observations, no visually distinct differences in overall disparity performance are noticeable between the various methods. This makes it difficult to draw firm conclusions based on visual analysis alone. Therefore, a rigorous assessment of the algorithm's performance requires a reliance on quantitative metrics, which provide the necessary data to compare the effectiveness of each approach, particularly in handling matching at object edges and across the entire image.

Table 6: Initialized difference map of different methods

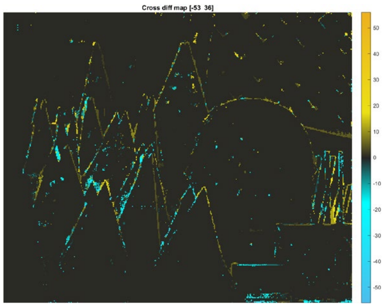
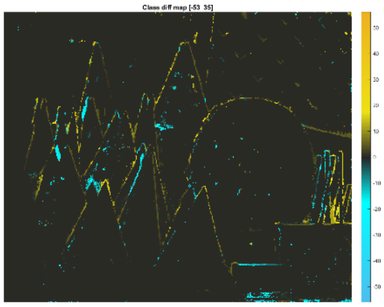
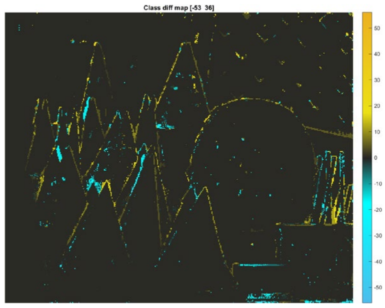
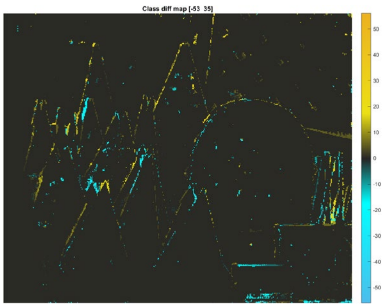
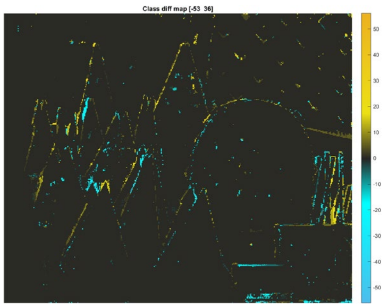
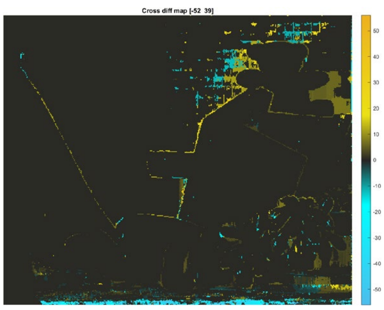
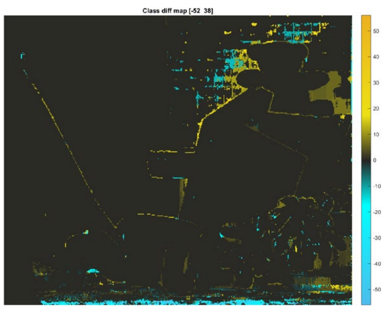
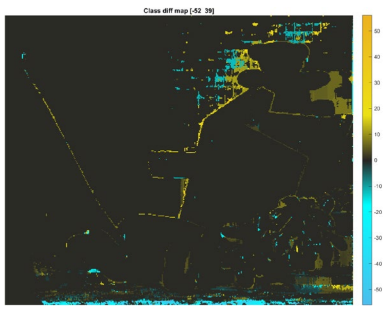
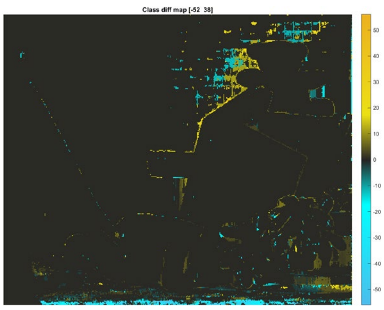
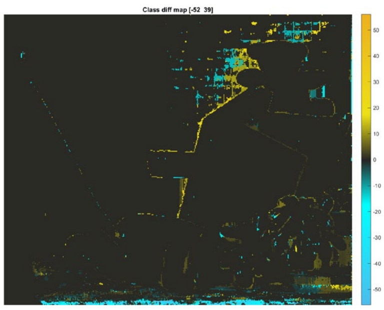
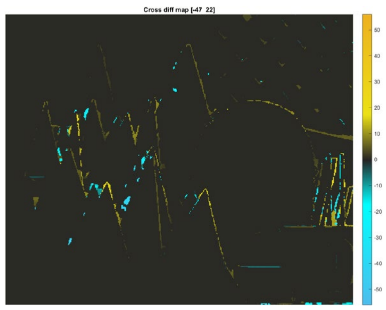
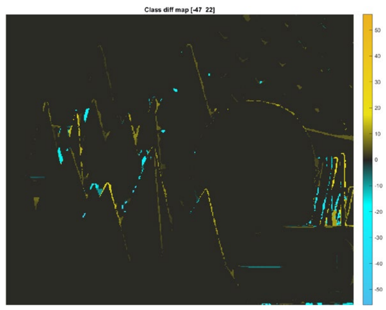
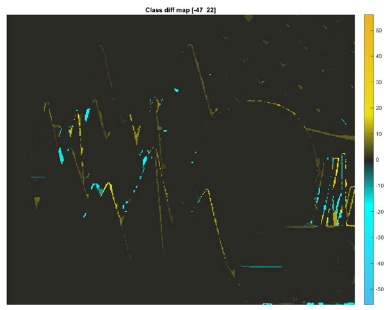
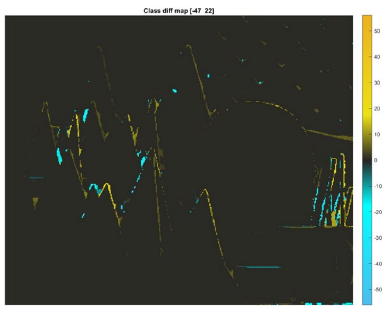
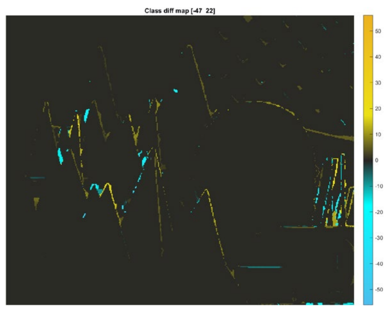
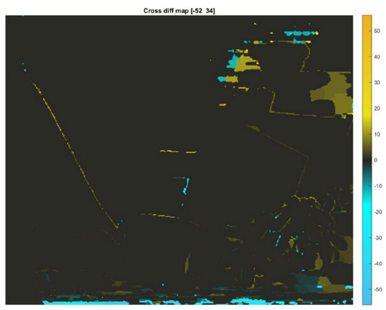
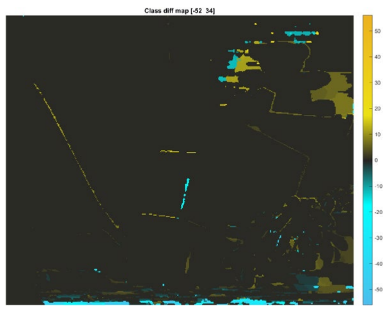
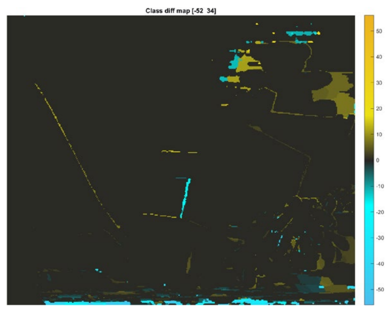
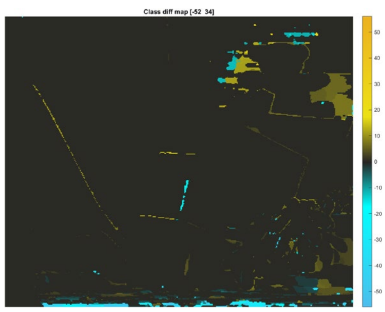
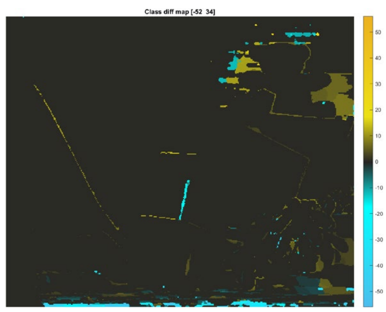
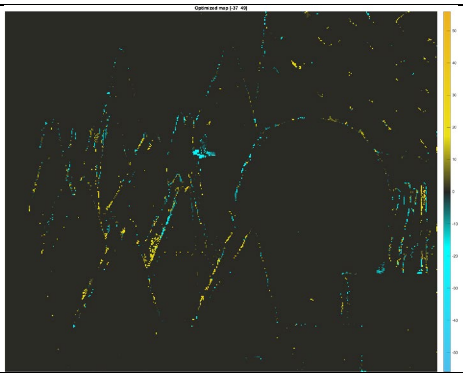
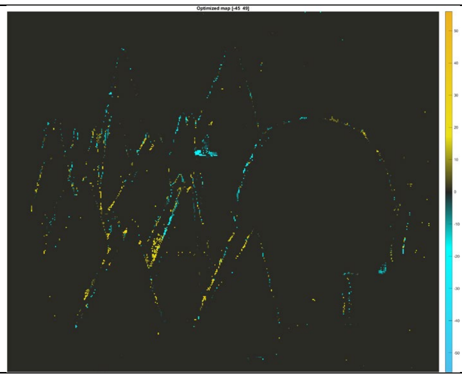
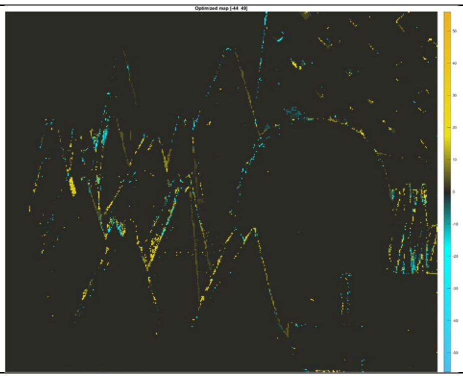
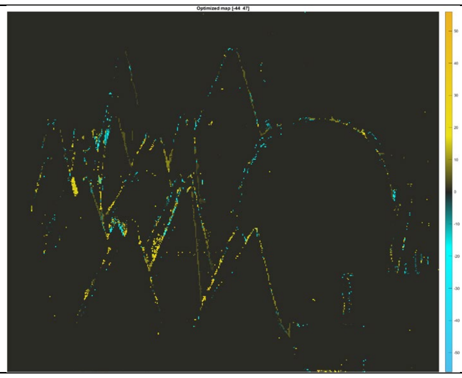
<table>
<tr><th colspan="6">Cones</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3">
</td><td colspan="3">
</td></tr>
<tr><th colspan="6">Teddy</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3">
</td><td colspan="3">
</td></tr>
</table>

Table 7: Final difference map of different methods

<table>
<tr><th colspan="6">Cones</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3">
</td><td colspan="3">
</td></tr>
<tr><th colspan="6">Teddy</th></tr>
<tr><td colspan="2">Cross-based</td><td colspan="2">Class-aided</td><td colspan="2">Partial Class-aided</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td colspan="3">Ground truth Class-aided</td><td colspan="3">Partial Ground truth Class-aided</td></tr>
<tr><td colspan="3">
</td><td colspan="3">
</td></tr>
</table>

The analysis of the optimized disparity maps in Figures 8 and 9 reveals several points. The adjusted regions in the optimized disparity maps are primarily concentrated around the edges of classified objects. This is expected, as object boundaries often present significant challenges for stereo-matching algorithms due to texture, lighting, and occlusion variations. The optimization process has effectively focused on these critical areas, enhancing the accuracy of disparity estimation where it is most needed. The introduction of ground truth-based adjustments has visibly improved the performance of the disparity map. On the visual representation, regions marked in yellow indicate areas where corrections have been applied effectively. This highlights the benefits of using ground truth information to refine the disparity maps, as it allows for precise adjustments and enhances overall accuracy. The clear improvement in these regions underscores the importance of high-quality classification and accurate reference data in optimizing stereo-matching results. Overall, the optimized maps demonstrate that targeted adjustments, especially around object boundaries, can significantly enhance disparity map accuracy. The use of ground truth data further refines these adjustments, leading to improved results.

Table 8: Initialized optimized map of different methods

| **Cones** | |
|---|---|
| Class-aided | Partial Class-aided |
|  |  |
| Ground truth Class-aided | Partial Ground truth Class-aided |
|  |  |

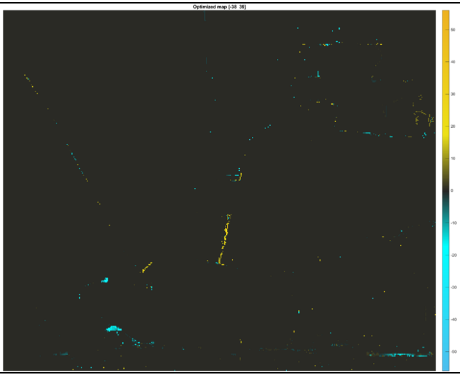
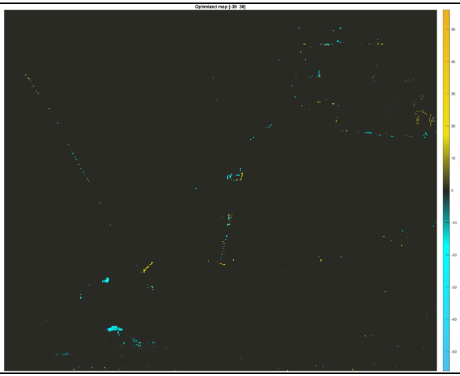
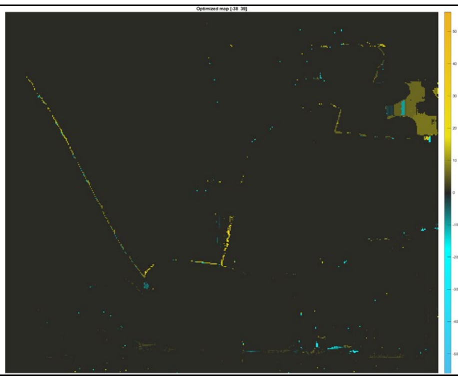
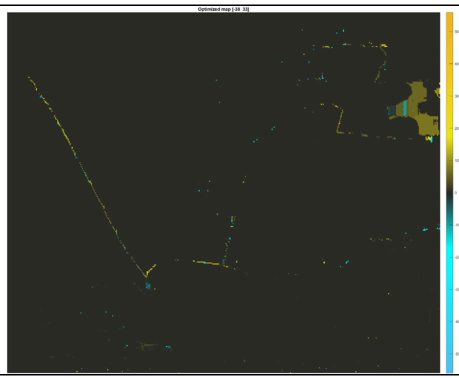
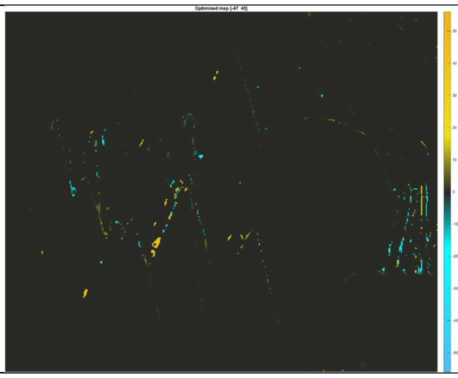
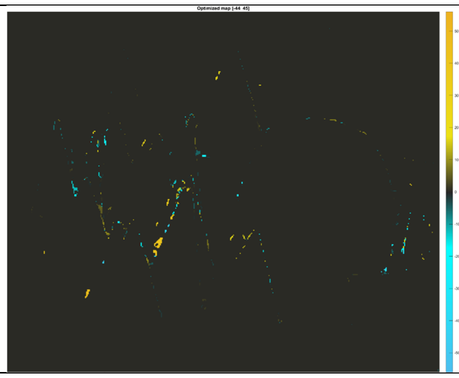
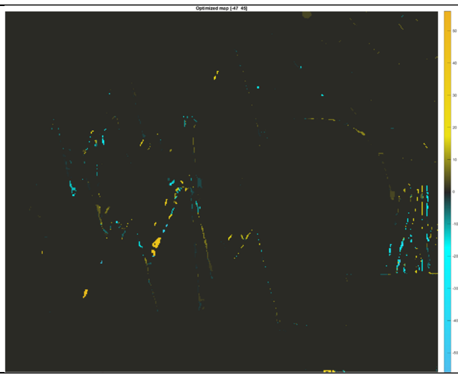
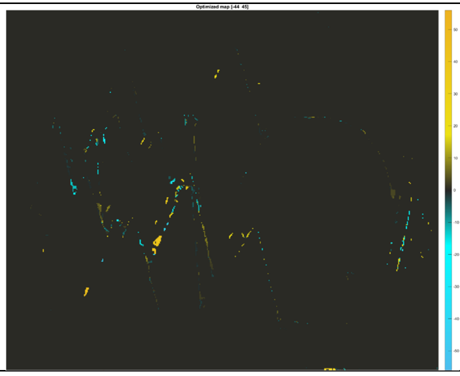
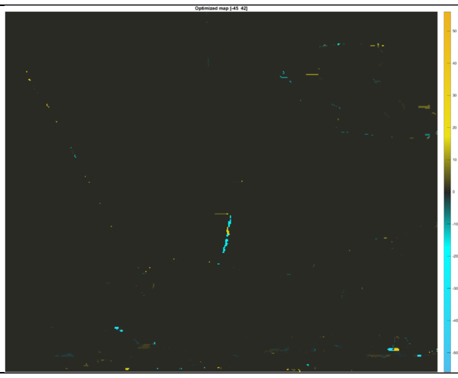
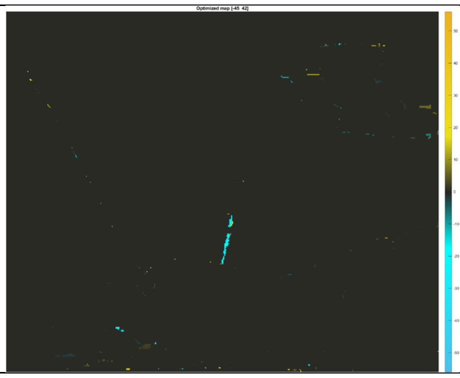
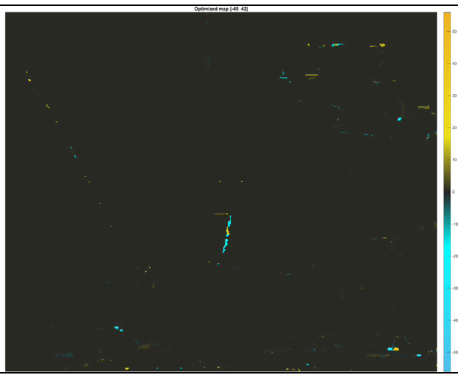
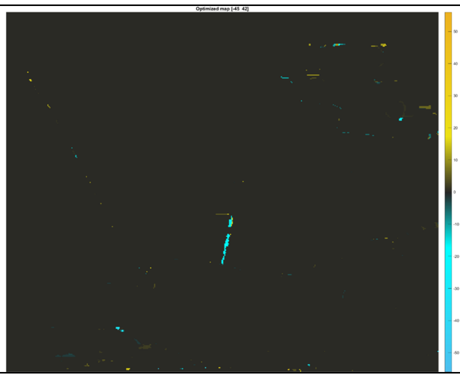
| **Teddy** | |
|---|---|
| Class-aided | Partial Class-aided |
|  |  |
| Ground truth Class-aided | Partial Ground truth Class-aided |
|  |  |

Table 9: Final optimized map of different methods

| **Cones** | |
|---|---|
| Class-aided | Partial Class-aided |
|  |  |

| Ground truth Class-aided | Partial Ground truth Class-aided |
|---|---|
|  |  |
| **Teddy** | |
| Class-aided | Partial Class-aided |
|  |  |
| Ground truth Class-aided | Partial Ground truth Class-aided |
|  |  |

Several key observations can be drawn regarding the impact of classification and ground truth on disparity map quality and the relationship between optimized area and error rate based on Figures 4 and 5:

- Impact of classification detailed on initialized disparities

The optimized area of Class-aided (8441) and Partial Class-aided (6094), and Ground truth Class-aided (21408) and Ground truth Partial Class-aided (15817) in Cones dataset, indicating that finer classification can provide better-initialized disparity images. Comparing the Whole Class-aided and Partial Class-aided approaches, the former generally performs better during initialization:

Classified images may initially produce more accurate disparity maps, however, the subsequent SGM may lead to a lower optimized area in the final disparity map. Thus, finer classification generally improves the initialized image but may require more refined adjustments during optimization.

- Effectiveness of accurate classification in reducing disparity range

The use of ground truth classification plays a critical role in refining the disparity maps, significantly improving results when applied to classified images. The data clearly shows the importance of accurate classification when comparing the impact of Class-aided and Ground Truth Class-aided, and Partial Class-aided and Partial Ground Truth Class-aided. The data shows that introducing accurate classification to generate the initialized disparity map could indeed increase the optimized area. Both Cones and Teddy datasets have seen this trend.

These findings highlight that accurate classification can significantly improve the approximation of disparity, especially when applied to classified images, emphasizing the importance of classification quality in stereo-matching accuracy.

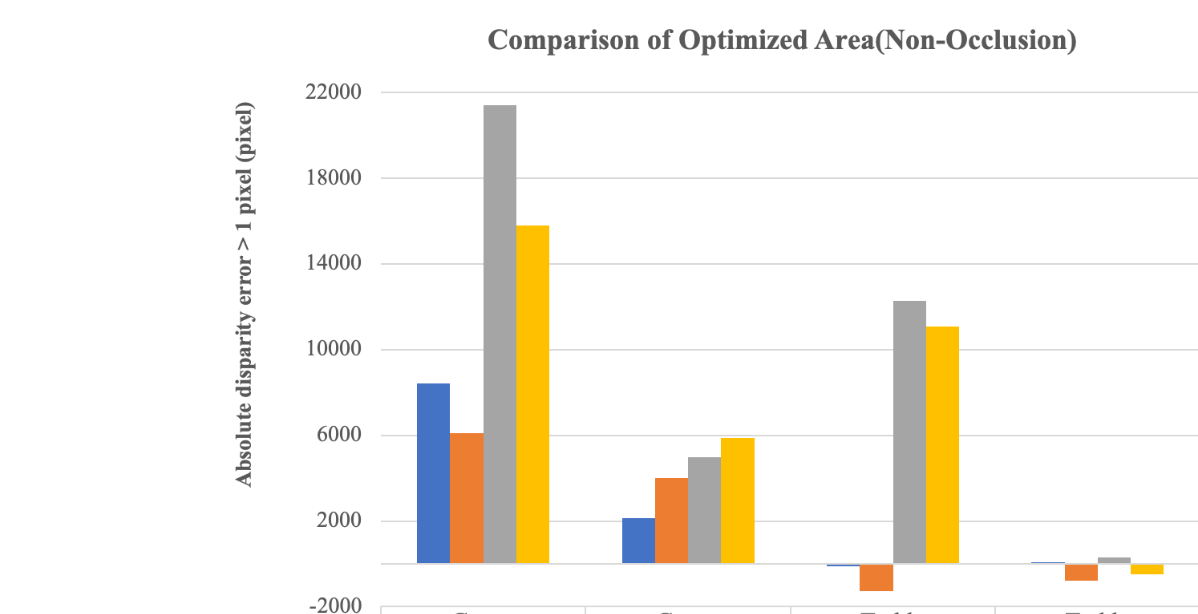
- Relationship between optimized area and error rate

Interestingly, the data reveals that a larger optimized area does not necessarily correlate with a lower error rate of matching results. For instance, Class-aided in the Cones dataset achieves a positive optimized area of 2153 pixels in the final map. Yet, its error rate (3.0210%) is not dramatically lower than that of Cross-based (2.9751%). The same trend can be observed in Partial Class-aided case. This shows that while a larger optimized area is generally desirable, it does not always guarantee a lower error rate. This discrepancy suggests that other factors, such as how well the optimization process handles specific challenging regions, e.g., edges or occlusions, play a crucial role in determining the final disparity accuracy.

- Impact of Class-aided Disparity on SGM

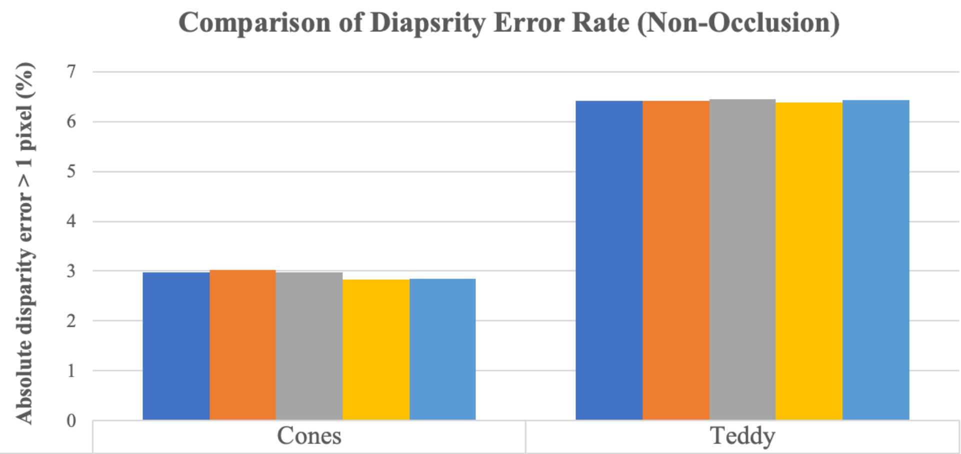
The Class-aided approach, by producing higher-quality initialized disparity maps, can assist in the automated decision-making of penalty parameters, making them better suited to the specific image set and more effective for subsequent SGM. This method has shown its effectiveness in certain datasets, as evidenced by positive values in the optimized area. However, the presence of negative values in other datasets indicates that even with higher-quality initialized disparity, errors may still occur during SGM, leading to less effective optimization compared to the initialized optimized area. Therefore, integrating classified information more effectively into the SGM stage remains a key area for future research.

In summary, finer classification improves the initial disparity map but may require more careful handling in the later SGM stages. Ground truth-based classification significantly enhances disparity map accuracy by constraining the disparity range, particularly in challenging areas like object boundaries. However, the relationship between the optimized area and error rate is not always straightforward. This highlights the importance of the algorithm's ability to effectively manage difficult regions. Overall, these findings emphasize the intricate nature of optimizing stereo-matching algorithms, where classification quality, accurate classification, and robust optimization techniques all play crucial roles in achieving the best results.




| | Cones (Initialized) | Cones (Final) | Teddy (Initialized) | Teddy (Final) |
|---|---|---|---|---|
| Class-aided | 8441 | 2153 | -116 | 66 |
| Partial Class-aided | 6094 | 4010 | -1278 | -791 |
| Ground truth Class-aided | 21408 | 4970 | 12289 | 307 |
| Partial Ground truth Class-aided | 15817 | 5897 | 11089 | -480 |

Figure 4: Optimized area of different methods




| | Cones | Teddy |
|---|---|---|
| Cross-based | 2.9751 | 6.4104 |
| Class-aided | 3.0210 | 6.4158 |
| Partial Class-aided | 2.9821 | 6.4476 |
| Ground truth Class-aided | 2.8285 | 6.3847 |
| Partial Ground truth Class-aided | 2.8480 | 6.4307 |

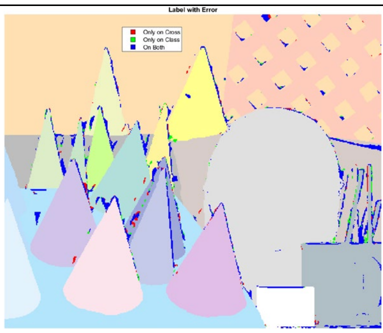
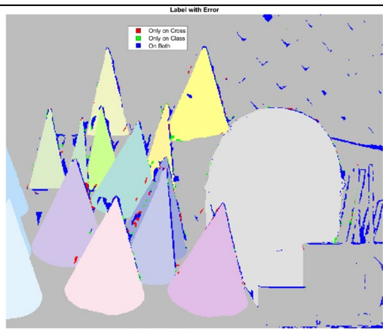
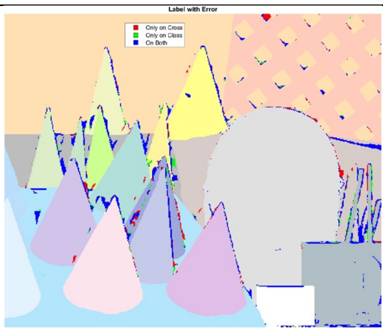
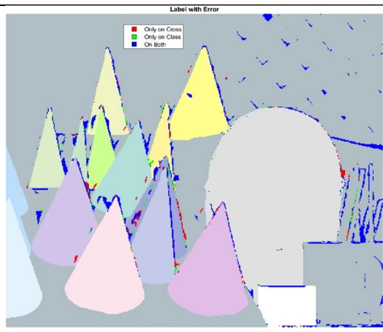
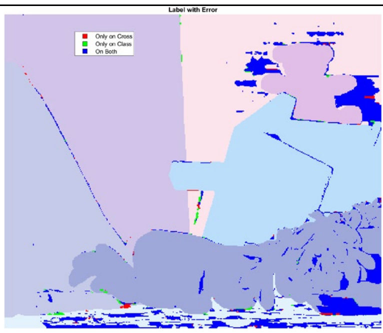
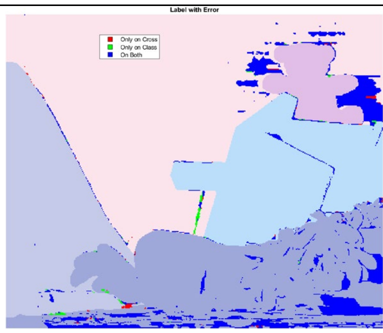
Figure 5: Error rate of different methods

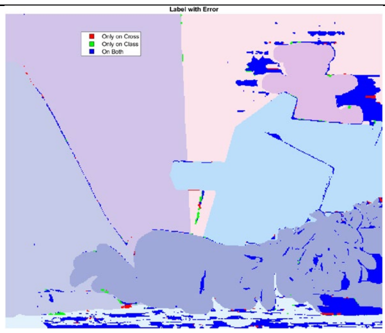
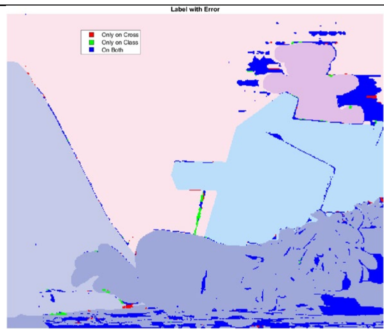
The error distribution map provides a spatial representation of disparity errors and the effectiveness of different methods. By marking pixels with three distinct colors—Red for error pixels corrected by the proposed methods, Green for error pixels present only in Cross-based matching, and Blue for error pixels occurring in both methods. The map visually highlights where improvements and persistent challenges lie. The use of these colors is significant as it provides a clear visual representation of the effectiveness of the methods, making it easier for the audience to interpret the map.

The error distribution map clearly shows that the majority of differences between the methods are concentrated around object borders. This aligns with the known challenge of accurately matching pixels at object boundaries due to occlusions, texture changes, and disparities in pixel correspondences.

The red pixels, representing errors corrected by the proposed methods, indicate that the Class-aided matching successfully refines disparity estimation, particularly by reducing errors near object borders. These corrections validate the proposed method’s effectiveness in improving depth accuracy by incorporating pixel classification constraints. However, the presence of green pixels, which represent errors unique to the Cross-based matching method, suggests that despite its traditional strength in stereo matching, there are areas, especially around complex object borders, where the Cross-based method fails. The proposed method offers a significant improvement. Conversely, the blue pixels, indicating errors that persist in both methods, highlight the limitations of the proposed class-constrained approach. These uncorrected pixels serve as a reminder that while the proposed method enhances the results, it still struggles in areas where accurate classification or segmentation is difficult. Blue pixels around object borders suggest that segmentation inaccuracies or challenging texture regions may prevent the method from fully eliminating errors.

Table 10: Error distribution map

<table>
<tr><th colspan="2">Cones</th></tr>
<tr><td>Class-aided</td><td>Partial Class-aided</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Ground truth Class-aided</td><td>Partial Ground truth Class-aided</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><th colspan="2">Teddy</th></tr>
<tr><td>Class-aided</td><td>Partial Class-aided</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

| Ground truth Class-aided | Partial Ground truth Class-aided |
| --- | --- |
|  |  |

Based on the quality assessment, the Ground truth Class-aided matching method was selected to generate a high-quality 3D point cloud that incorporates object labels. A comparison between this labeled point cloud and the original, unclassified (Figure 6) version reveals that the introduction of labels transforms raw point cloud data into structured, meaningful point cloud information (Figure 7). This labeling process significantly enhances the interpretability of the data by providing users with categorized, context-specific insights about the scene. The advantages of this transformation are particularly notable in applications such as surveying and mapping, where users can directly extract relevant information, such as object boundaries, classifications, and spatial relationships, from the labeled point cloud. This enriched dataset enables more detailed and accurate environmental analysis, making it highly actionable for decision-making processes.

This study's primary contribution lies in showcasing how the integration of object classification within point clouds bridges the gap between unstructured data and actionable information. The result is a more refined and informative representation, offering valuable insights for future applications, particularly in fields like geospatial analysis and remote sensing applications.

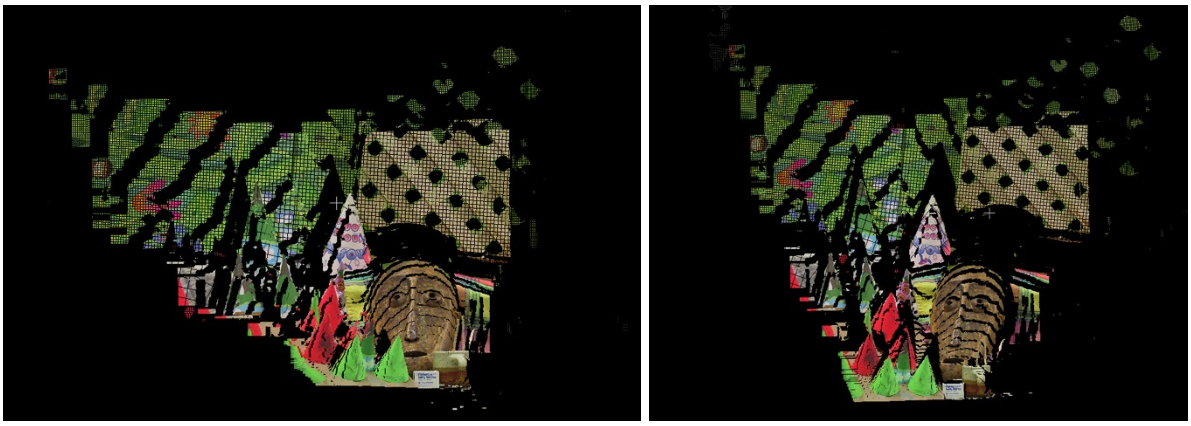


Figure 6: Point Cloud data

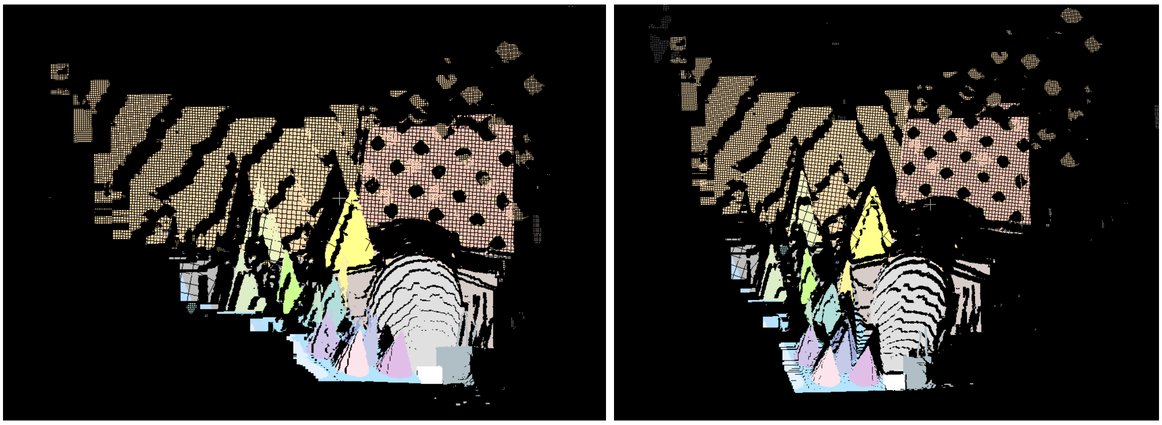


Figure 7: Point Cloud information

**Conclusion and Recommendation**

The results of this study underscore the significant benefits of utilizing higher-quality classified images to enhance the refinement of disparity map generation. Despite these advantages, several technical and operational constraints currently limit the practical feasibility of this approach:

1. Challenges with Current Classification Models

   Existing classification models often lack the accuracy required for effective object classification within the proposed method. Although these models can provide preliminary classifications, the extensive manual corrections needed afterward undermine the efficiency of the algorithm, rendering it labor-intensive and time-consuming.

2. Limitations of Absolute Constraints

   The absolute constraints initially imposed by the algorithm are not sustainable with current technological capabilities. In scenarios where perfectly accurate classified images cannot be obtained, the method must adapt by using more flexible, relative constraints during the matching process. This adjustment will help address inaccuracies introduced by imperfect classifications and improve overall performance.

In this study, the pixel costs generated from Cross-based and Class-aided approaches are integrated into the subsequent SGM process to determine the overall matching cost. When the initialized disparity is produced during the first stage, the constraint of assigning infinite cost values to pixel pairs with mismatched labels between the left and right images is carried forward into the SGM stage. This means that if a pixel in the right image falls outside the valid label range within the disparity scope, it will be assigned a prohibitively high cost. As a result, under SGM's winner-takes-all principle, such disparity values are excluded from being selected. This mechanism underscores the importance of high-quality initialized disparity maps. The constraint imposed by the pixel cost ensures that mismatched labels in the right image are effectively penalized, thus improving the robustness of the disparity estimation process. Incorporating this cost structure highlights how classification information aids in narrowing the search range and minimizing errors, further emphasizing the role of label consistency in stereo-matching accuracy.

The successful creation of a labeled 3D point cloud using the Ground truth Class-aided matching method underscores the broader impact of this approach. By transforming raw point cloud data into structured, labeled information, this method has significant implications for applications such as surveying, mapping, and geospatial analysis, where enhanced interpretability and decision-making are essential.

Future work should concentrate on optimizing the integration of classified information into the SGM process. This entails developing methods that balance improved classification accuracy with the minimization of new errors, thus enhancing the robustness and efficiency of disparity map algorithms. Additionally, research should explore dynamic or adaptive constraints to address challenges related to object boundaries and complex textures.

In summary, while this study illustrates the potential of using classification-based constraints to enhance disparity map algorithms, ongoing research is necessary to refine current classification techniques and integration methods. Advancements in these areas could lead to more efficient and automated algorithms, effectively leveraging available data and reducing the need for manual interventions, thereby maximizing the potential of classified information.

**References**

Hamzah, R. A., Abd Rahim, R., & Noh, Z. M. (2010). Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application. *2010 3rd International Conference on Computer Science and Information Technology.*

Hirschmuller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2*, 807-814.

Hong, L., & Chen, G. (2004). Segment-based stereo matching using graph cuts. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* (Vol. 1, pp. I-I). IEEE.

Pessoa, G. G., Amorim, A., Galo, M., & Galo, M. d. L. B. T. (2019). Photogrammetric point cloud classification based on geometric and radiometric data integration. *Boletim de Ciências Geodésicas*, *25*.

Scharstein, D., & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.

Ting, H.W. (2014). The Optimization of Integrating Cross-based Matching into Semi-Global Matching. Thesis, National Taiwan University, pp178.

Ting, H. W., & Jaw, J. J. (2017). Integrating Cross-based Matching into Semi-Global Matching. *Journal of Photogrammetry and Remote Sensing, 22*(3), 157-180.

Yang, W., Sheng, S., Luo, X., & Xie, S. (2022). Geometric relation based point clouds classification and segmentation. *Concurrency and Computation: Practice and Experience*, *34*(11), e6845.

Yang, Q., Wang, L., Yang, R., Wang, S., Liao, M., & Nister, D. (2006). Real-time Global Stereo Matching Using Hierarchical Belief Propagation. In *Bmvc* (Vol. 6, pp. 989-998).

Yeh, H. C. (2012). Hardware Components for Real-Time Stereo Matching: Acceleration of 3D HD TV with FPGAs. In. Delft University of Technology.

Yoo, J.-C., & Han, T. H. (2009). Fast normalized cross-correlation. *Circuits, systems and signal processing, 28*, 819-843.

Zhang, K., Lu, J., & Lafruit, G. (2009). Cross-based local stereo matching using orthogonal integral images. *IEEE transactions on circuits and systems for video technology, 19*(7), 1073-1079.