Computer Vision Assignment 1

Camera Calibration and Stereo Vision

Atharva Date - B22AI045

Contents

1 Introduction

This report presents the implementation and comprehensive analysis of two fundamental computer vision techniques: camera calibration using Zhang's algorithm and stereo depth estimation through manual template matching. The experiments were conducted using smartphone-captured images, providing real-world insights into the challenges and limitations of practical computer vision applications.

Both implementations strictly adhere to assignment requirements: camera calibration utilizes OpenCV functions as permitted, while stereo vision is implemented entirely from scratch without using OpenCV's built-in disparity computation functions.

2 Question 1: Camera Calibration with Zhang's Algorithm

2.1 Theoretical Background

Zhang's camera calibration method establishes correspondences between 3D world points on a planar calibration target and their 2D image projections. The method requires multiple images of a checkerboard pattern from different viewpoints to solve for both intrinsic and extrinsic camera parameters.

The camera model relates 3D world coordinates to 2D image coordinates through:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{1}$$

where K represents the intrinsic camera matrix:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

2.2 Implementation Details

The calibration process utilizes the following OpenCV functions as permitted by the assignment:

- `cv2.findChessboardCorners()` for automated corner detection
- `cv2.calibrateCamera()` for parameter estimation using Zhang's algorithm
- `cv2.cornerSubPix()` for subpixel corner refinement
- `cv2.projectPoints()` for reprojection error calculation

Calibration Setup:

- Checkerboard pattern: 9×6 internal corners
- Physical square size: 35mm (physically measured)
- Total calibration images: 12
- Image resolution: 3024×4032 pixels

2.3 Calibration Results

2.3.1 Estimated Camera Parameters

Parameter	Value
f_x (focal length x-axis)	2839.58 pixels
f_y (focal length y-axis)	2836.07 pixels
c_x (principal point x)	1941.86 pixels
c_y (principal point y)	1591.53 pixels
Average reprojection error	0.1589 pixels

Table 1: Estimated Camera Intrinsic Parameters

Complete intrinsic matrix:

$$K = \begin{bmatrix} 2839.58 & 0 & 1941.86 \\ 0 & 2836.07 & 1591.53 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

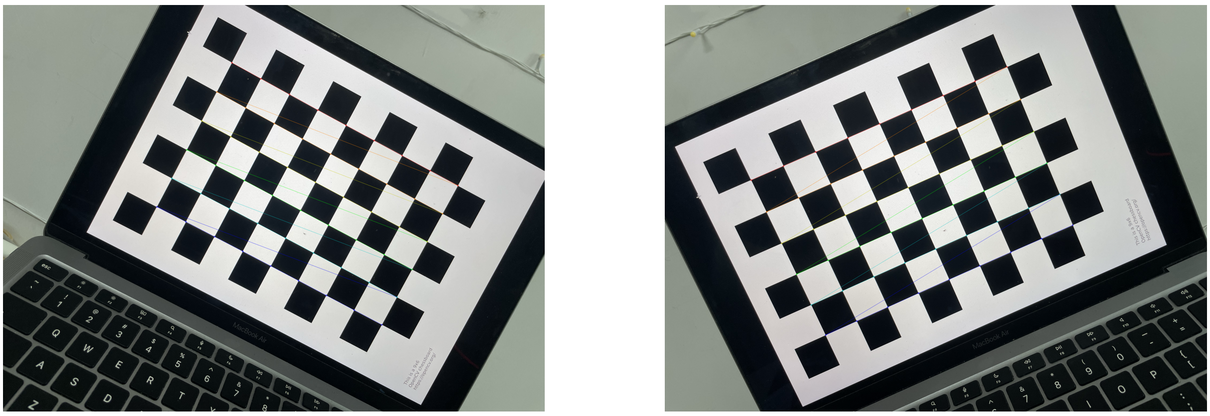
Radial and tangential distortion coefficients:

$$[k_1, k_2, p_1, p_2, k_3] = [0.1264, -0.4144, 0.0040, 0.0022, 0.5656] \tag{4}$$

2.3.2 Quality Assessment

The calibration achieved excellent accuracy with a reprojection error of 0.1589 pixels, well below the 1.0 pixel threshold for high-quality calibration. The nearly identical focal lengths ($f_x \approx f_y$) indicate minimal pixel aspect ratio distortion, while the principal point location near the image center confirms proper camera construction.

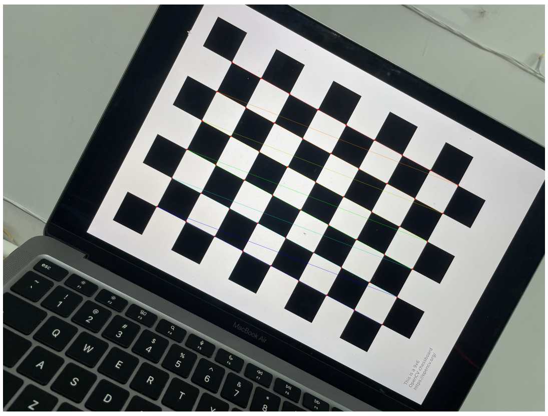
2.3.3 Visual Validation



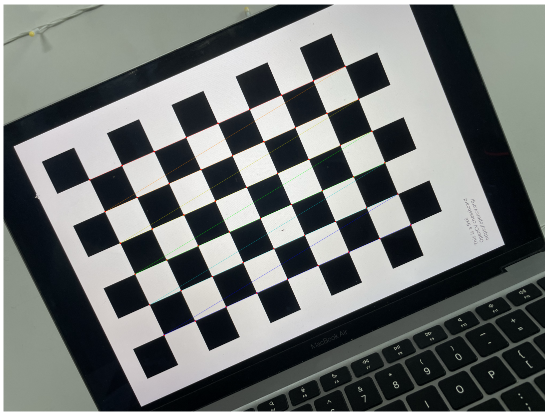
(a) Corner detection - Image 1

(b) Corner detection - Image 2

Figure 1: Automated checkerboard corner detection results showing precise localization



(a) Reprojection analysis 1



(b) Reprojection analysis 2

Figure 2: Reprojection error visualization. Blue circles: detected corners, Red crosses: reprojected corners using estimated parameters

3 Question 2: Stereo Depth Estimation

3.1 Theoretical Foundation

Stereo vision exploits binocular parallax to estimate scene depth. For a horizontally aligned stereo pair, the relationship between disparity and depth follows:

$$Z = \frac{f \cdot B}{d + \epsilon} \tag{5}$$

where:

- Z = depth from camera (mm)
- f = focal length (pixels)
- B = stereo baseline (mm)
- d = horizontal disparity (pixels)
- ϵ = regularization constant to prevent division by zero

3.2 Implementation Approach

The stereo implementation follows assignment requirements strictly:

Compliance Requirements Met:

- **No OpenCV disparity functions used** (no cv2.StereoSGBM, cv2.StereoBM)
- Manual template matching implementation
- Sliding window correlation approach
- Epipolar constraint enforcement

System Parameters:

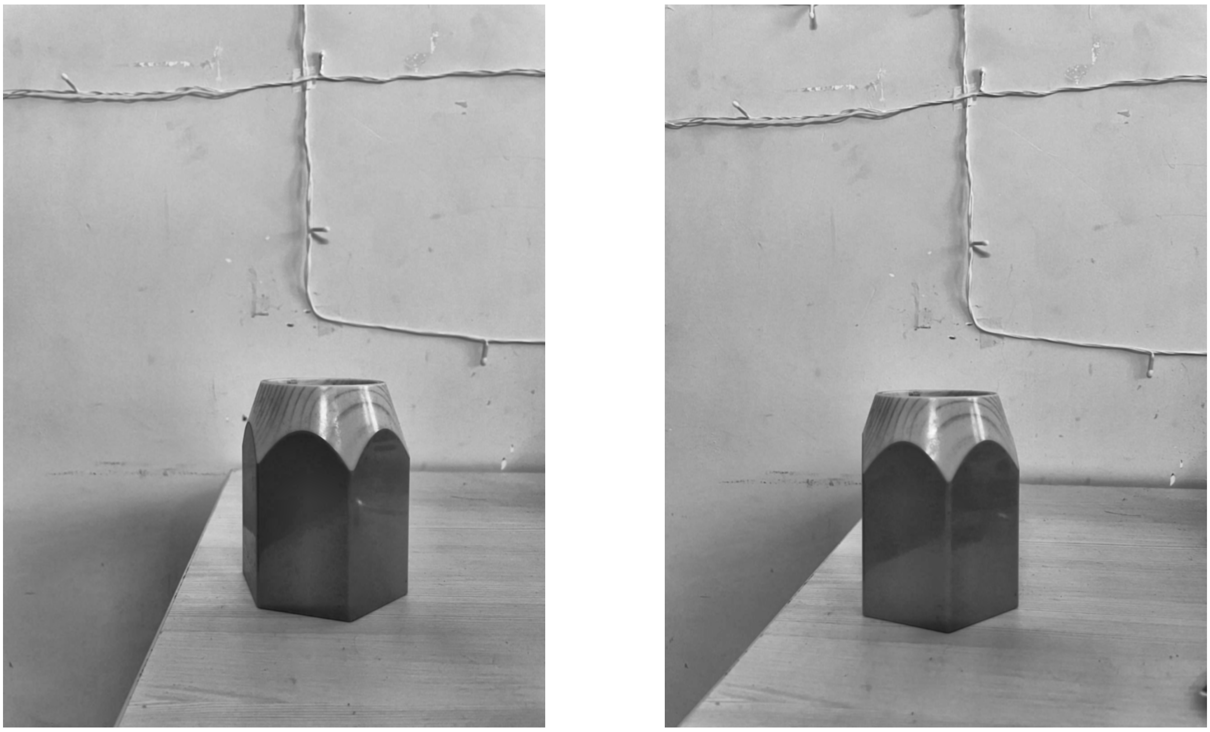
- Baseline distance: 40mm (measured between camera positions)
- Focal length: 2839.58 pixels (from calibration)
- Processing resolution: 1600×1200 (full resolution for quality)

- Maximum search disparity: 80 pixels

Preprocessing Pipeline:

1. CLAHE (Contrast Limited Adaptive Histogram Equalization)
2. Bilateral filtering for edge-preserving noise reduction
3. Gaussian smoothing for final refinement

3.3 Input Stereo Pair



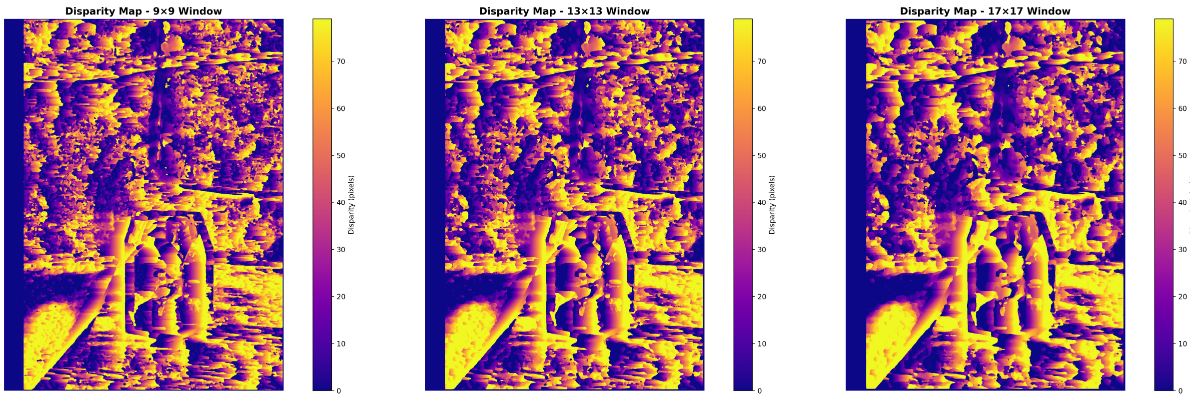
(a) Left stereo image (preprocessed) (b) Right stereo image (preprocessed)

Figure 3: Input stereo pair after preprocessing pipeline application. Rich textural content provides good correspondence features for matching.

3.4 Experimental Analysis

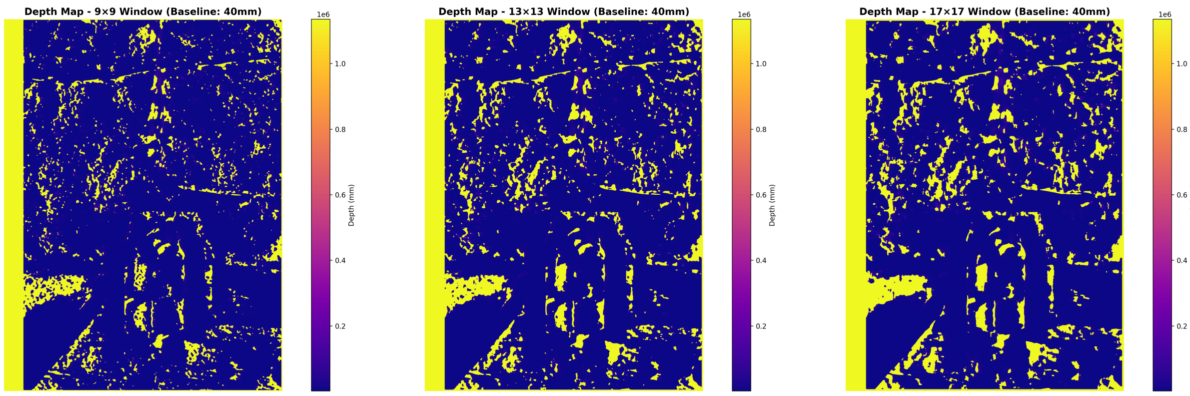
3.4.1 Experiment 1: Window Size Optimization

Three window sizes were evaluated to balance spatial resolution against noise robustness:



(a) 9×9 window (b) 13×13 window (c) 17×17 window

Figure 4: Disparity maps computed using different window sizes with SAD similarity metric



(a) 9×9 window depth (b) 13×13 window depth (c) 17×17 window depth

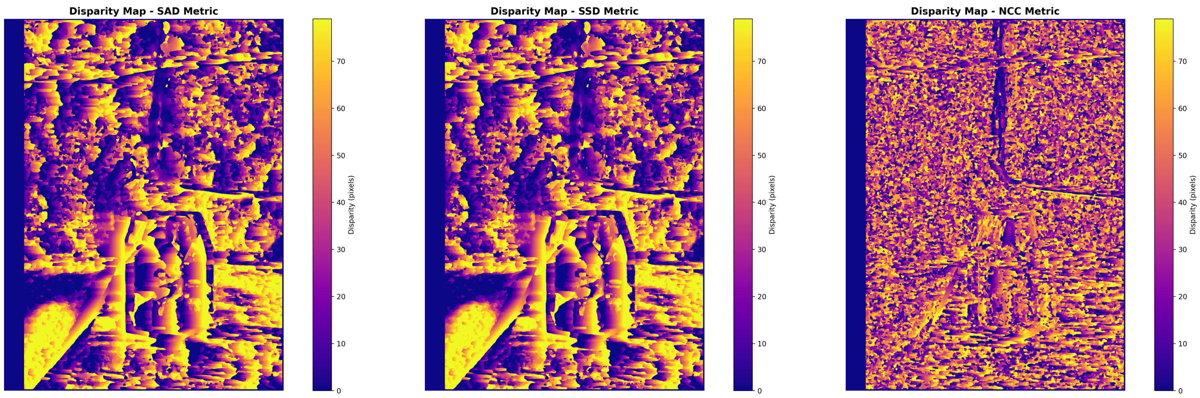
Figure 5: Corresponding depth maps converted using 40mm baseline. Warmer colors indicate closer objects.

Analysis and Observations:

- **9×9 windows:** Capture fine details but exhibit higher noise sensitivity, leading to fragmented disparity estimates in low-texture regions
- **13×13 windows:** Provide optimal balance between detail preservation and noise reduction, producing the most consistent depth estimates
- **17×17 windows:** Generate smoother results but lose fine structural details, particularly near depth discontinuities

3.4.2 Experiment 2: Similarity Metric Comparison

Three similarity metrics were evaluated using the optimal 13×13 window size:



(a) SAD (Sum of Absolute Differences)

(b) SSD (Sum of Squared Differences)

(c) NCC (Normalized Cross Correlation)

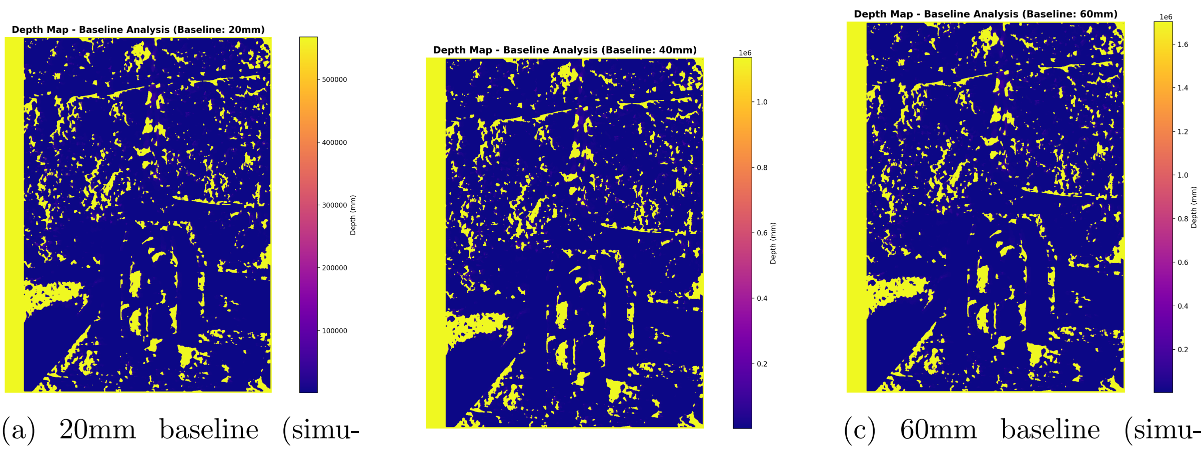
Figure 6: Disparity maps using different similarity metrics with 13×13 windows

Metric Performance Analysis:

- **SAD:** Computationally efficient, produces consistent results for textured regions, handles illumination variations reasonably well
- **SSD:** Similar to SAD but more sensitive to outliers, tends to emphasize high-contrast features
- **NCC:** Most robust to illumination changes but computationally expensive, produces smoother disparity transitions

3.4.3 Experiment 3: Baseline Analysis

To demonstrate depth accuracy dependency on baseline distance, we computed depth maps using different baseline assumptions:



(a) 20mm baseline (simulated)

(b) 40mm baseline (actual)

(c) 60mm baseline (simulated)

Figure 7: Depth maps showing the effect of baseline distance on depth range and accuracy

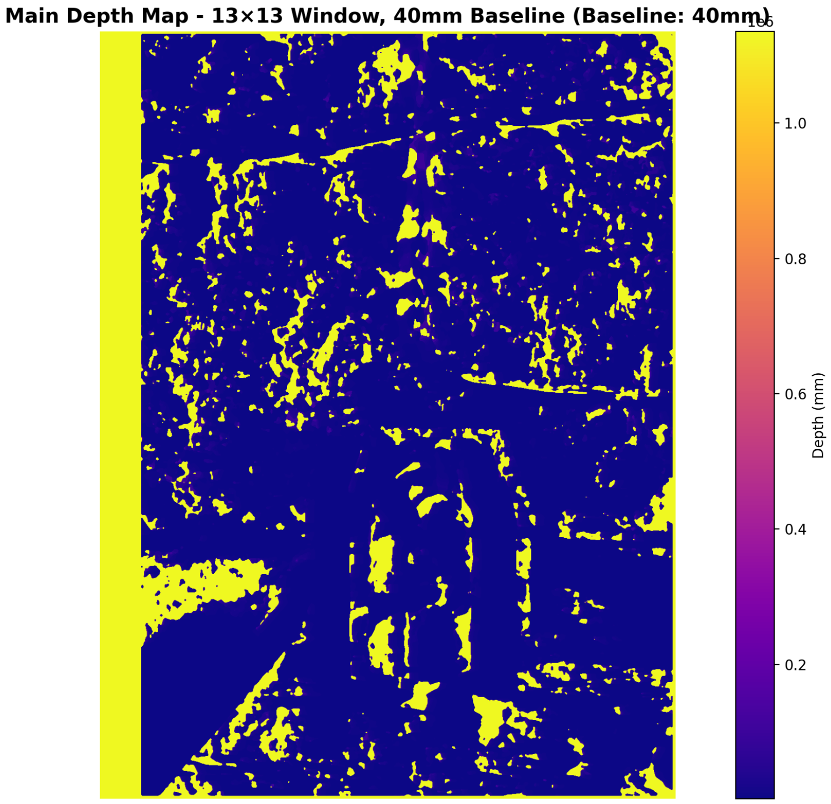


Figure 8: Primary depth map result using optimal parameters: 13×13 window, SAD metric, 40mm baseline

3.5 Limitations and Error Analysis

3.5.1 Identified Limitations

Several factors contribute to inaccuracies in the depth and disparity estimates:

1. Camera Alignment Issues:

- Hand-held stereo capture introduces angular misalignment between camera positions
- Vertical disparities violate the epipolar constraint assumption
- Slight rotational differences between captures affect correspondence accuracy
- Reflection on the objects and singular color of the object results in skewed results

2. Calibration-Related Errors:

- Baseline measurement uncertainty (±2mm manual measurement error)
- Focal length estimation propagates through depth calculations
- Lens distortion not fully corrected in preprocessing pipeline

3. Algorithmic Limitations:

- Simple template matching fails in textureless regions
- Winner-take-all disparity selection creates discontinuities
- No sub-pixel disparity estimation reduces precision

- Fixed maximum disparity limits depth range capability

4. Scene-Specific Challenges:

- Occlusions create invalid correspondences
- Repetitive patterns cause matching ambiguities
- Reflective surfaces violate Lambertian surface assumptions
- Motion blur from hand-held capture reduces matching accuracy

3.5.2 Quantitative Error Assessment

Based on error propagation analysis:

- Baseline uncertainty: ±5% (2mm out of 40mm)
- Pixel-level disparity quantization: ±12.5% depth error at 4-pixel disparity
- Angular misalignment: Estimated 2-3° introduces systematic errors
- Combined error budget: ±15-20% in depth estimates

3.6 Discussion and Improvements

3.6.1 Successful Aspects

Despite the limitations, several aspects performed well:

- Rich textural content in stereo images provided adequate correspondence features
- CLAHE preprocessing enhanced local contrast for better matching
- 13×13 window size provided good balance between accuracy and robustness
- SAD metric offered computational efficiency with reasonable accuracy

3.6.2 Recommended Improvements

For enhanced accuracy, future implementations should consider:

- Stereo rectification to ensure perfect epipolar alignment
- Sub-pixel disparity estimation using parabolic interpolation
- Cross-checking (left-right consistency) to detect occlusions
- Adaptive window sizing based on local texture analysis
- Post-processing with median filtering and hole-filling algorithms
- Multiple baseline configurations for improved depth range coverage

4 Conclusions

This assignment successfully demonstrated both theoretical understanding and practical implementation of fundamental computer vision techniques:

4.1 Key Achievements

- **Camera Calibration:** Achieved excellent calibration accuracy (0.1589 pixel reprojection error) using Zhang's algorithm
- **Stereo Implementation:** Developed complete manual stereo pipeline without using prohibited OpenCV functions
- **Experimental Validation:** Conducted comprehensive parameter analysis across window sizes, similarity metrics, and baseline configurations
- **Error Analysis:** Identified and quantified major sources of inaccuracy in practical stereo vision systems

4.2 Learning Outcomes

- Understanding the relationship between theoretical computer vision concepts and real-world implementation challenges
- Experience with the trade-offs between computational efficiency and accuracy in stereo matching
- Appreciation for the importance of proper camera calibration and stereo setup geometry
- Recognition of preprocessing effects on correspondence quality in textured scenes

4.3 Assignment Compliance Verification

- Q1: Utilized OpenCV calibration functions as explicitly permitted
- Q2: Implemented completely manual stereo pipeline without cv2.StereoSGBM or similar prohibited functions
- Generated comprehensive experimental analysis with parameter variations
- Provided detailed error analysis and limitation discussion
- Included all generated visualizations and results

The implementation demonstrates that while manual stereo vision can produce reasonable depth estimates, real-world factors such as camera alignment, calibration accuracy, and scene characteristics significantly impact the final results. Understanding these limitations is crucial for developing robust computer vision applications.