DepthSense™: Accurate Distance Measurement for Assistive Navigation

Authors: Patel, A., Chen, S., Kim, D., Ramirez, L., & Johnson, M.
Date: 2024
Institution: iVision AI Research Division
Contact: research@ivisionai.org

Abstract

This paper presents DepthSense™, a novel multi-modal distance measurement system designed specifically for assistive navigation applications. By combining hardware-based depth sensing, structure-from-motion algorithms, and learned object dimension models, DepthSense™ achieves unprecedented accuracy in smartphone-based distance estimation. Our extensive testing demonstrates ±5cm accuracy for objects within 3 meters and ±12cm accuracy for objects 3-10 meters away. User studies with 52 blind participants show significant improvements in navigation confidence and obstacle avoidance compared to existing assistive technologies. DepthSense™ represents a significant advancement in spatial awareness tools for visually impaired individuals, enabling more confident and independent navigation in unfamiliar environments.

1. Introduction

Accurate spatial awareness is fundamental to confident navigation for visually impaired individuals. While object recognition technologies have advanced significantly in recent years, precise distance measurement remains a critical challenge for assistive navigation systems. Current smartphone-based solutions typically achieve accuracies of only ±15-20cm, which is insufficient for safe navigation in complex environments.

DepthSense™ addresses this gap through a comprehensive approach to distance estimation that leverages multiple sensing modalities and advanced computational techniques. Our system is designed to:

1. Provide accurate distance measurements on consumer smartphone hardware
2. Function effectively across diverse environmental conditions
3. Integrate seamlessly with object recognition systems
4. Deliver real-time performance suitable for navigation assistance

This paper details the technical architecture of DepthSense™, our evaluation methodology, and the system's performance in both controlled testing and real-world user studies.

2. Related Work

2.1 Smartphone-Based Distance Estimation

Smartphone-based distance estimation has evolved significantly with the introduction of dedicated depth sensors in premium devices. Keselman et al. (2017) evaluated early LiDAR implementations on mobile devices, reporting accuracies of ±3cm but only at close range (< 1 meter). More recent work by Zhang et al. (2022) demonstrated improved performance with newer hardware, achieving ±8cm accuracy up to 5 meters.

For devices without dedicated depth sensors, monocular depth estimation techniques have been explored. Cao et al. (2021) proposed a deep learning approach for monocular depth estimation on smartphones, achieving ±12cm accuracy for objects within 3 meters. However, their system required significant computational resources and performed poorly in low-light conditions.

2.2 Computer Vision Approaches

Vision-based distance estimation typically relies on either stereoscopic vision or monocular cues. Wang and Shen (2020) demonstrated a stereoscopic approach using smartphone camera movement to create a baseline for triangulation, achieving ±10cm accuracy under ideal conditions, but with significant performance degradation in complex environments.

Structure-from-motion (SfM) techniques have shown promise for mobile applications. Liu et al. (2023) implemented a real-time SfM system for smartphones that achieved ±7cm accuracy for static scenes, though performance degraded significantly with moving objects or rapid camera movement.

2.3 Assistive Technology Applications

Within assistive technology, distance estimation has received increasing attention. Mascetti et al. (2021) integrated basic distance estimation into an object recognition system for blind users, but reported accuracy of only ±25cm, which users found insufficient for confident navigation.

Rahman et al. (2022) developed a dedicated obstacle detection and distance estimation system for white cane users, achieving ±15cm accuracy for obstacles at ground level, but their system did not address objects at variable heights or provide a complete spatial mapping of the environment.

3. System Design

DepthSense™ employs a multi-modal approach to distance estimation, combining hardware sensors, computer vision algorithms, and learned object properties. Figure 1 illustrates the system architecture.

3.1 Core Components

DepthSense™ consists of four primary components:

3.1.1 Hardware Depth Sensing

On devices equipped with ToF (Time of Flight) sensors or LiDAR, DepthSense™ directly incorporates hardware depth measurements. The system accounts for known error patterns in these sensors through a calibration model trained on our extensive depth reference dataset. This component provides:

- Direct depth measurements within the sensor's range (typically 0-5 meters)
- Confidence scores for each measurement based on sensor characteristics
- Error correction based on environmental factors (lighting, reflectivity)

3.1.2 Structure-from-Motion Engine

For devices without dedicated depth sensors, or to extend the range of hardware sensors, DepthSense™ implements a real-time Structure-from-Motion (SfM) algorithm. This component:

- Tracks distinctive features across frames during natural camera movement
- Reconstructs 3D positions of these features through triangulation
- Creates sparse depth maps that are refined over time
- Maintains a sliding window of recent frames to optimize for mobile performance

Our SfM implementation is based on a modified ORB-SLAM2 architecture (Mur-Artal and Tardós, 2017), optimized for real-time performance on mobile devices through selective feature processing and GPU acceleration.

3.1.3 Object Dimension Database

DepthSense™ incorporates a comprehensive database of typical dimensions for over 2,000 common objects. This allows distance estimation based on apparent object size when an object is recognized with high confidence. The database includes:

- Average dimensions for common objects (furniture, vehicles, household items)
- Variance models to account for size variations within object categories
- Distinctive dimension ratios to resolve ambiguity in partial object views

The system employs Bayesian inference to combine size-based distance estimates with other measurement modalities, weighting each source according to its reliability in the current context.

3.1.4 Spatial Consistency Engine

The Spatial Consistency Engine enforces geometric constraints between detected objects and environmental features. This component:

- Validates distance estimates against known physical constraints
- Infers relative positions between objects
- Detects and corrects outlier measurements
- Builds a consistent 3D scene model over time

This approach significantly improves accuracy by leveraging the physical coherence of the environment, particularly in complex scenes with multiple objects.

3.2 Integration and Fusion Algorithm

The core innovation of DepthSense™ lies in its adaptive fusion algorithm that combines inputs from all available modalities. Rather than using a fixed weighting scheme, the system employs a dynamic confidence-based integration approach.

For each detected object, the fusion algorithm:

1. Collects distance estimates from all available modalities
2. Assigns confidence scores to each estimate based on:
    - Sensor reliability in current conditions
    - Consistency with previous measurements
    - Agreement with other modalities
    - Environmental factors (lighting, motion, etc.)
3. Generates a weighted estimate that minimizes expected error
4. Tracks estimate stability over time to reduce jitter

Figure 2 illustrates this fusion process in different measurement scenarios.

The adaptive nature of this approach allows DepthSense™ to maintain high accuracy across diverse hardware platforms and environmental conditions.

4. Implementation

4.1 Software Architecture

DepthSense™ is implemented as a modular C++ library with lightweight bindings for iOS and Android platforms. The implementation prioritizes:

- Real-time performance on mid-range smartphone hardware
- Minimal memory and battery impact

- Thread-safe operation for integration with other systems
- Fallback strategies for different hardware capabilities

The software stack includes specialized optimization for common mobile SoCs, including neural processing units on recent Qualcomm Snapdragon, Apple A-series, and Google Tensor chips.

4.2 Integration with Object Recognition

DepthSense™ is designed to work seamlessly with object recognition systems, particularly our DenseVision™ framework. The integration provides:

- Shared feature extraction to reduce computational overhead
- Object instance segmentation for precise distance measurement
- Tracking of objects across frames for temporal consistency
- Priority-based processing focusing on navigation-relevant objects

This tight integration enables the complete spatial mapping required for effective navigational guidance.

4.3 Optimization Techniques

Several key optimizations enable real-time performance on mobile devices:

1. **Adaptive sampling:** Processing resolution varies based on scene complexity
2. **Region-based processing:** Computational resources focus on regions containing detected objects
3. **Temporal coherence:** Information from previous frames reduces per-frame computation
4. **Hardware acceleration:** Leverages GPU/NPU for neural network components and feature extraction

5. **Progressive refinement:** Provides initial estimates quickly, then refines with additional processing

These optimizations allow DepthSense™ to operate at 10+ frames per second on mid-range smartphones while consuming approximately 15% of the CPU and 20% of the GPU resources.

5. Evaluation

We conducted comprehensive evaluations of DepthSense™ through both technical benchmarking and user studies.

5.1 Technical Evaluation

5.1.1 Methodology

The technical evaluation employed a rigorous methodology:

1. **Reference environment:** Testing in a controlled environment with laser-measured ground truth distances
2. **Diverse scenarios:** 12 environmental scenarios including indoor residential, office, retail, and outdoor urban settings
3. **Variable conditions:** Testing across different lighting conditions, device movements, and object arrangements
4. **Hardware variation:** Evaluation on 8 smartphone models representing different hardware capabilities
5. **Comparison systems:** Benchmarking against 4 commercial and 2 research-based distance estimation systems

For each test case, we measured absolute error, relative error, and processing time.

5.1.2 Results

Accuracy Performance:

Table 1 summarizes accuracy results across distance ranges.

Table 1: Distance Estimation Accuracy by Range

Distance Range	Mean Absolute Error	90th Percentile Error	Standard Deviation
0-1 meters	±2.3 cm	±4.1 cm	1.7 cm
1-3 meters	±4.8 cm	±7.6 cm	2.9 cm
3-5 meters	±9.2 cm	±14.3 cm	5.2 cm
5-10 meters	±16.7 cm	±24.5 cm	8.4 cm

Figure 3 compares DepthSense™ accuracy against alternative systems.

Environmental Factors:

Table 2 shows performance across different environmental conditions.

Table 2: Impact of Environmental Factors on Accuracy

Condition	Mean Absolute Error	Performance Degradation
Optimal lighting	±4.2 cm	Baseline
Low light	±6.8 cm	62%

Bright/direct light	±7.3 cm	74%
Moving camera	±5.9 cm	40%
Reflective surfaces	±8.4 cm	100%
Transparent objects	±11.2 cm	167%

Hardware Variation:

Figure 4 illustrates performance across different device categories.

Computational Performance:

Table 3 presents computational metrics on representative devices.

Table 3: Computational Performance Metrics

Device Category	Frames Per Second	CPU Usage	GPU Usage	Battery Impact
Flagship	18.7 FPS	12%	18%	2.8% per hour
Mid-range	11.3 FPS	17%	24%	3.9% per hour
Budget	7.2 FPS	26%	31%	5.2% per hour

5.2 User Study

5.2.1 Participants and Methodology

We conducted a comprehensive user study with 52 blind participants (29 female, 23 male) aged 21-73 (mean = 46.2, SD = 13.7). Participants represented diverse backgrounds in terms of:

- Onset of blindness: 33 congenital, 19 acquired
- Mobility aid usage: 38 white cane users, 14 guide dog users
- Technology proficiency: 20 high, 25 medium, 7 low (self-reported)

The study employed a within-subjects design where participants completed navigation tasks both with DepthSense™-enhanced guidance and with a baseline system providing object recognition without accurate distance information. Tasks included:

1. Navigating an unfamiliar room and locating specific objects
2. Traversing a path with obstacles of various heights
3. Maintaining appropriate distance from other pedestrians in a public space

For each task, we measured completion time, navigation errors, collisions, and subjective confidence ratings.

5.2.2 Results

Task Performance:

Table 4 summarizes task performance metrics.

Table 4: Navigation Task Performance Metrics

Metric	With DepthSense™	Baseline System	Improvement
Task completion time	73.4s	126.8s	42.1%
Navigation errors	1.3	3.9	66.7%
Obstacle collisions	0.7	2.8	75.0%
Path deviation distance	0.85m	1.72m	50.6%

User Experience:

Participants reported significantly higher confidence when using DepthSense™ (mean rating 8.3/10) compared to the baseline system (mean rating 5.6/10), $t(51) = 9.73$, $p < .001$.

Figure 5 illustrates subjective ratings across different aspects of the navigation experience.

Qualitative Findings:

Thematic analysis of interview data revealed several key themes:

1. **Spatial Confidence:** 47/52 participants reported increased confidence in their understanding of the environment
2. **Proactive Navigation:** 42/52 participants noted the ability to plan movements in advance rather than reacting to encountered obstacles
3. **Distance Precision:** 39/52 participants specifically valued the precise distance information for maintaining appropriate social distance and finding objects
4. **Learning Curve:** 28/52 participants mentioned a brief learning period (5-10 minutes) to adapt to the distance-aware guidance
5. **System Trust:** 36/52 participants reported developing trust in the system after successful navigation experiences

Representative quotes include:

"Having accurate distances completely changes how I move through a space. Instead of cautious shuffling, I can walk with purpose." - P17, 42, congenitally blind

"The difference between knowing something is 'nearby' versus knowing it's exactly 2 meters ahead is profound for my confidence." - P34, 29, acquired blindness

"I could maintain appropriate distance from others at the cafe counter, which has always been a source of anxiety for me." - P08, 56, congenitally blind

6. Discussion

6.1 Key Contributions

DepthSense™ makes several significant contributions to the field of assistive technology:

1. **Multi-modal fusion approach:** By combining hardware sensors, computer vision, and object knowledge, DepthSense™ achieves superior accuracy across diverse conditions
2. **Mobile optimization:** Real-time performance on consumer smartphones makes the technology widely accessible
3. **Environmental adaptability:** Consistent performance across varied lighting conditions and environments enables practical everyday use
4. **User-centered metrics:** Evaluation focused not just on technical accuracy, but on navigational outcomes that matter to visually impaired

users

The significant improvements in navigation metrics demonstrate that accurate distance information fundamentally changes how blind users can interact with their environment, enabling more confident and efficient movement.

6.2 Limitations

Despite its strong performance, DepthSense™ has several limitations:

1. **Hardware dependency:** While the system functions on all smartphones, optimal performance requires devices with dedicated depth sensors
2. **Environmental challenges:** Performance degrades with highly reflective surfaces, transparent objects, and extremely bright lighting
3. **Moving objects:** Current implementations have reduced accuracy for rapidly moving objects
4. **Range limitations:** Accuracy diminishes significantly beyond 10 meters
5. **Battery impact:** Continuous use has a noticeable impact