CSSR-Nav: Culturally Sensitive Social Robot Navigation Using Open-Source Vision-Language Model and Empirical Cultural Knowledge

Birhanu Shimelis Girma[1], Ibrahim Jimoh[1], Assane Gueye[1], Moise Busogi[1], and David Vernon[1]

Carnegie Mellon University Africa, Kigali, Rwanda
{bgirmash, ioj, assaneg, mbusogi}@andrew.cmu.edu, david@vernon.eu

Abstract. We present CSSR-Nav (Culturally Sensitive Social Robot Navigation), a novel vision-language model based navigation approach that enables social robots to navigate in accordance with culture-specific social norms. Unlike existing social navigation methods that rely on generic Western etiquette or require extensive training datasets, our approach leverages empirical cultural knowledge derived from a survey of 143 Rwandan respondents and an open-source vision-language model (LLaVA) for real-time, edge-computed decision making. We introduce a novel *Rwandan Social Costmap Layer* that integrates directly with the ROS navigation stack, encoding cultural norms such as maintaining appropriate personal space (1 meter), avoiding interruption of conversations, passing behind social groups, and showing respect to elders. Our system runs entirely on edge hardware (Jetson Orin Nano) mounted on a Pepper humanoid robot, enabling deployment in resource-constrained environments without cloud dependencies. Through systematic ablation studies with 10 experimental conditions and 50+ trials, we demonstrate that CSSR-Nav achieves significant improvements in cultural compliance while maintaining real-time performance. This work represents the first integration of empirically-derived African cultural norms into autonomous social robot navigation.

Keywords: Social Robot Navigation · Vision-Language Models · Cultural Robotics · Human-Robot Interaction · Costmap Layers

1 Introduction

Social robots are increasingly deployed in human-centric environments such as hospitals, hotels, museums, and educational institutions [1–5]. For these robots to be effective and accepted, they must navigate not only safely but also in ways that respect the social norms and cultural expectations of the people they serve [6]. However, the vast majority of social navigation research has been conducted in Western contexts, with social compliance defined according to Western etiquette [7].

Recent taxonomies in social navigation research have established fundamental properties that characterize socially aware robotic agents [8]. A robot demonstrates social awareness when it: (1) recognizes human agents as distinct entities deserving prioritized safety considerations; (2) operates in ways that minimize disruption, discomfort, and confusion for nearby people; (3) communicates its navigational intentions, whether through explicit signals or implicit behavioral cues; and (4) when faced with spatial conflicts, evaluates the social context and adopts resolution strategies that may prioritize human comfort over task efficiency. While these properties provide a valuable framework for social navigation, they assume culturally-neutral definitions of "discomfort" and "social manner" assumptions we challenge in this work.

The challenge of cultural specificity in social navigation: Social norms vary significantly across cultures [9]. Behaviors considered polite in one culture may be perceived as rude or inappropriate in another. For example, the appropriate interpersonal distance, eye contact patterns, and passing behaviors differ substantially between cultures [10]. Yet, current social navigation systems either encode generic rules (*e.g.*"pass on the right") or learn from datasets collected predominantly in Western environments [11, 12], limiting their applicability to diverse global contexts.

The gap in African social robotics: Despite growing interest in deploying social robots across Africa for applications ranging from healthcare to education [13], there exists no navigation system specifically designed to respect African cultural norms. The CSSR4Africa (Culturally Sensitive Social Robotics for Africa) project addresses this gap by first establishing what constitutes culturally appropriate behavior through empirical research [13].

Limitations of existing VLM-based approaches: Recent work has demonstrated the potential of Vision-Language Models (VLMs) for social navigation [14]. However, these approaches rely on proprietary cloud-based models (*e.g.*GPT-4V), which present challenges for deployment in regions with limited internet connectivity, raise data privacy concerns, and incur ongoing operational costs. Furthermore, they encode social norms through prompts based on assumed universal etiquette rather than empirically validated cultural knowledge.

Main Contributions. In this paper, we present CSSR-Nav, a culturally sensitive social navigation system that addresses these limitations. Our main contributions include:

1. **Empirically-grounded cultural navigation:** This work is the first to integrate empirically-derived cultural norms (from a survey of 143 Rwandan respondents) into an autonomous social navigation system, encoding 7 specific cultural behaviors including personal space maintenance, conversation non-interruption, and elder respect protocols.
2. **Novel Rwandan Social Costmap Layer:** We introduce a ROS navigation stack plugin that translates cultural norms into costmap modifications, enabling standard path planners to generate culturally-appropriate trajectories without algorithm changes.

3. **Edge-deployed open-source VLM:** We demonstrate real-time social perception using LLaVA on a Jetson Orin Nano, achieving practical inference times without cloud dependencies, making our system suitable for deployment in resource-constrained environments.
4. **Comprehensive ablation study:** We conduct systematic ablation tests with 10 experimental conditions to quantify the individual and combined effects of cultural norms on navigation performance, social compliance, and user acceptance.

2 Related Work

2.1 Social Robot Navigation

Social navigation extends traditional robot navigation by considering humans not merely as dynamic obstacles but as social entities with expectations about robot behavior [1, 15]. Research has addressed various aspects including personal space [16], passing behavior [17], and group awareness [18].

Learning-based approaches have gained prominence, with imitation learning [19, 20] and reinforcement learning [21, 22] showing promise. However, these methods require extensive training data and may not generalize to contexts different from their training distribution. The SCAND dataset [11] and MuSoHu [12] provide valuable resources but were collected exclusively in Western environments.

2.2 Vision-Language Models for Navigation

VLMs have emerged as powerful tools for robotic decision-making due to their contextual understanding and commonsense reasoning capabilities [**?**]. LM-Nav [23] demonstrated outdoor navigation using GPT-3 and CLIP. VLM-Social-Nav [14] applied GPT-4V to social navigation, achieving significant improvements over baseline methods.

However, reliance on proprietary cloud models limits practical deployment. Recent open-source VLMs such as LLaVA [24] offer comparable capabilities with the advantage of local deployment, though their application to social navigation remains unexplored.

2.3 Culture and Social Robotics

The influence of culture on human-robot interaction has been documented across dimensions including proxemics [25], communication style [26], and social role expectations [27]. However, this knowledge has rarely been operationalized in navigation systems.

The CSSR4Africa project [13] represents a systematic effort to establish cultural requirements for social robots in African contexts. This work builds directly on the Rwandan Cultural Knowledge survey [28], translating empirical findings into computational navigation behaviors.

Table 1: Rwandan Cultural Norms Relevant to Navigation

ID	Cultural Norm	Implementation
3-1	Maintain distance of one meter or less when passing	Personal space radius
2-26	Do not walk between conversing people	Group interaction cost
3-3	Pass behind groups of people	Directional preference
2-27	Do not walk ahead of elders	Elder respect zone
2-11	Approach to greet, do not wave from distance	Greeting approach
2-18	Begin interactions with courteous greeting	Greeting trigger
2-9	Bow slightly when greeting	Respectful gesture

3 Rwandan Cultural Knowledge for Navigation

This approach is grounded in empirical cultural knowledge derived from the CSSR4Africa Rwandan Cultural Knowledge Survey [28]. This survey collected responses from 143 participants at Carnegie Mellon University Africa, establishing consensus on 57 cultural behaviors relevant to social robot interaction.

3.1 Navigation-Relevant Cultural Norms

From the survey results, multiple norms are identified with direct implications for robot navigation. Table 1 summarizes the key norms and their implementation in the system.

3.2 Formalizing Cultural Costs

Following the formulation in [14], we define the navigation cost function as:

$$\mathcal{C}(\mathbf{s}, \mathbf{a}) = \alpha \cdot \mathcal{C}_{\text{goal}} + \beta \cdot \mathcal{C}_{\text{obst}} + \gamma \cdot \mathcal{C}_{\text{cultural}} \tag{1}$$

where $\mathcal{C}_{\text{goal}}$ encourages movement toward the goal, $\mathcal{C}_{\text{obst}}$ discourages collisions, and $\mathcal{C}_{\text{cultural}}$ encourages adherence to cultural norms. Unlike previous work that defines $\mathcal{C}_{\text{social}}$ generically, we decompose $\mathcal{C}_{\text{cultural}}$ into specific, empirically-grounded components:

$$\mathcal{C}_{\text{cultural}} = \mathcal{C}_{\text{space}} + \mathcal{C}_{\text{group}} + \mathcal{C}_{\text{elder}} + \mathcal{C}_{\text{direction}} \tag{2}$$

Each component corresponds to a specific cultural norm: $\mathcal{C}_{\text{space}}$ for personal space maintenance (Norm 3-1), $\mathcal{C}_{\text{group}}$ for conversation non-interruption (Norm 2-26), $\mathcal{C}_{\text{elder}}$ for elder respect positioning (Norm 2-27), and $\mathcal{C}_{\text{direction}}$ for pass-behind preference (Norm 3-3).

4 System Architecture

Fig. 2 illustrates the CSSR-Nav system architecture. The approach integrates three key components: LLaVA-based social perception, the Rwandan Social Costmap Layer, and culture-aware behavior coordination.

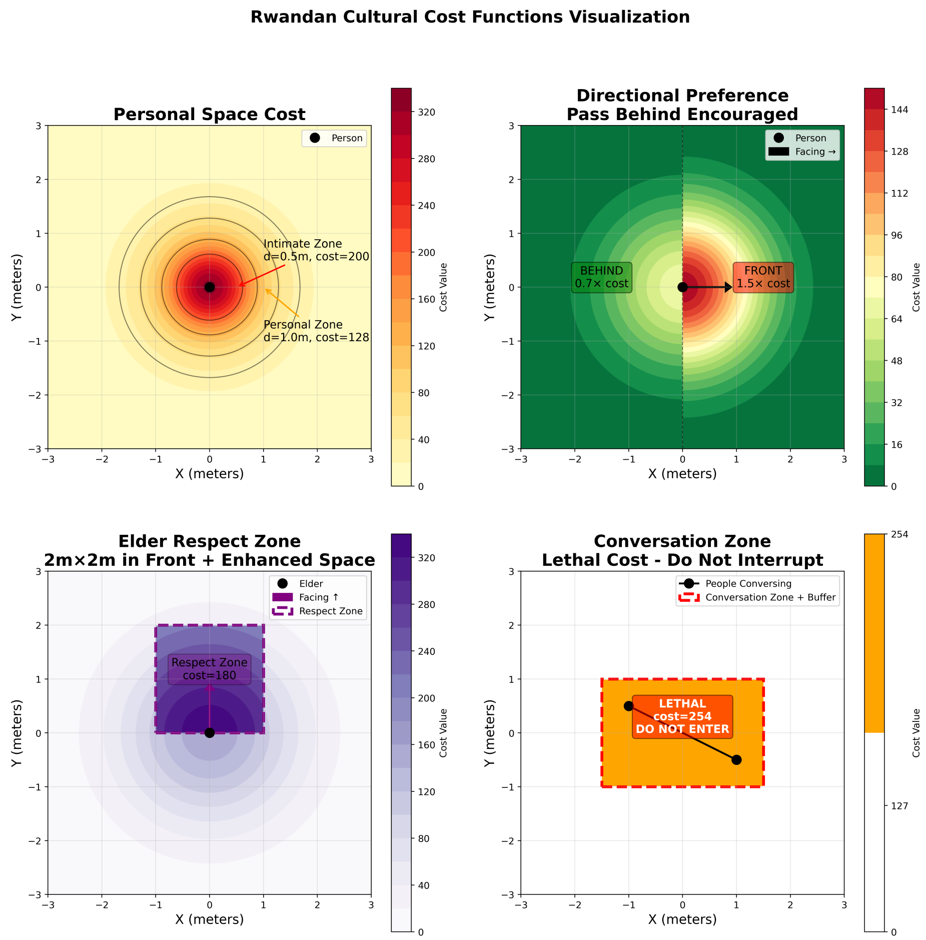

Fig. 1: Rwandan Social Costmap Layer integrates with ROS costmap_2d through standard layer interfaces (initialize, updateBounds, updateCosts) and implements four cultural cost functions: personal space, directional preference, elder respect zones, and conversation interruption avoidance.

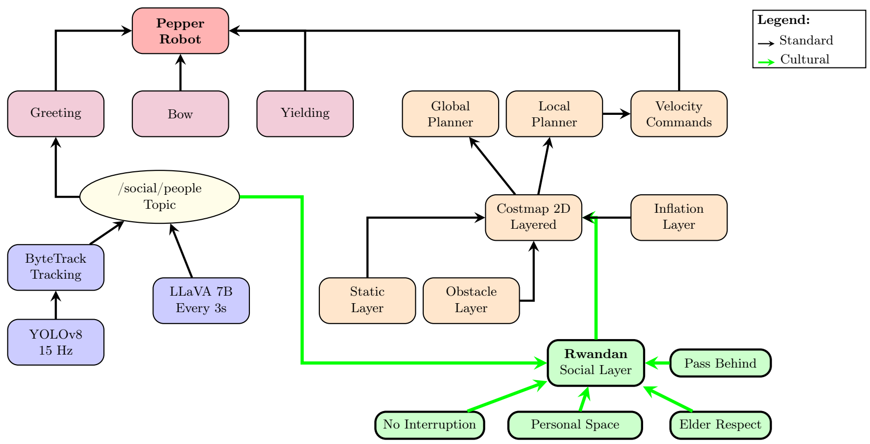

Fig. 2: System architecture showing the integration of perception, navigation (Costmap 2D with Rwandan Social Layer), and culturally-aware behavior coordination.

4.1 Hardware Platform

Our system is deployed on a SoftBank Pepper robot augmented with: Intel RealSense D435 camera providing color images at 640×480 resolution and depth information for social perception; YDLidar G4, 2D LiDAR for obstacle detection and localization; and NVIDIA Jetson Orin Nano, Edge computing platform running perception and VLM inference.

This configuration enables fully edge-deployed processing without cloud dependencies, addressing connectivity limitations common in many deployment environments. Table 2 details the hardware specifications.

Table 2: Hardware Platform Specifications

Component	**Model**	**Specifications**
Robot Base	Pepper (SoftBank)	Height: 1.21m, Weight: 28kg
RGB-D Camera	RealSense D435	640×480 @ 30fps, Depth range: 0.3-3m
LiDAR	YDLidar G4	360°, 10m range, 5-12Hz
Compute	Jetson Orin Nano	8GB RAM, 1024-core GPU, 6-core ARM CPU

4.2 LLaVA-based Social Perception

The system employs LLaVA (Large Language and Vision Assistant) [24] for social scene understanding. Unlike VLM-Social-Nav which uses GPT-4V through cloud API, we run LLaVA locally on the Jetson Orin Nano using 4-bit quantization for memory efficiency.

Perception Pipeline: To achieve real-time performance, the system adopts a two-stage perception approach: (1) Fast Detection (YOLOv8): Runs continuously at 15 Hz to detect people, providing real-time tracking without VLM latency; and (2) Social Understanding (LLaVA): Triggered every 3 seconds when social entities are detected, querying the VLM for deeper social context understanding including age estimation, group detection, and activity recognition.

Algorithm 1 details the perception pipeline that fuses fast detection with periodic VLM queries.

Algorithm 1 Two-Stage Social Perception Pipeline

Require: RGB-D frame I_t, depth frame D_t
Ensure: Person detections $\mathcal{P}_t$, social groups $\mathcal{G}_t$

```
P_t ← YOLOv8(I_t) {Fast detection at 15Hz}
Track persons across frames using IoU matching
if t − t_last_vlm > Δt_vlm AND |P_t| > 0 then
    S ← LLaVA(I_t, cultural_prompt) {VLM query}
    Extract age groups, facing directions from S
    G_t ← DetectGroups(P_t, S)
    t_last_vlm ← t
end if
Estimate 3D positions using depth D_t and camera intrinsics
return P_t, G_t
```

Culturally-Adapted Prompting: A key innovation of this approach is the integration of Rwandan cultural norms directly into VLM prompts. Fig. 3 shows our prompt structure, which differs from generic social navigation prompts by explicitly encoding surveyed cultural expectations.

Input Image: [RGB frame from RealSense or pepper builtin camera]
System Prompt: You are the vision system for a social robot navigating in Carnegie Mellon University Africas AI and Robotics lab while respecting Rwandan social norms. Analyze this scene following Rwandan cultural norms.
Cultural Guidelines:

– Maintain at least 1 meter distance when passing
– Never walk between people who are conversing
– Pass behind groups, not in front
– Do not walk ahead of elderly persons
– Approach people to greet; do not wave from distance

Output Format (JSON only): {"people_count": N, "positions": [x, y], "groups_conversing": bool, "elder_present": bool, "clear_path": "left/center/right"}

Fig. 3: Culturally-adapted prompt for LLaVA. Unlike generic prompts, we explicitly encode Rwandan cultural norms derived from the CSSR4Africa survey.

4.3 Rwandan Social Costmap Layer

The core technical contribution of this work is the Rwandan Social Costmap Layer, a plugin for the ROS navigation stack that translates cultural norms into costmap modifications. Algorithm 2 presents the costmap update procedure.

Algorithm 2 Rwandan Social Costmap Update

Require: People $\mathcal{P}$, Groups $\mathcal{G}$, Costmap C
Ensure: Updated costmap C'
$C' \leftarrow C$ {Initialize with base costmap}
for each person $p \in \mathcal{P}$ **do**
  Apply personal space cost (Eq. 3) around p
  if p is elder **then**
    Apply elder respect zone cost (Eq. 7) ahead of p
    Multiply personal space by $\mu_{\text{elder}} = 1.2$
  end if
  Apply directional preference cost (Eq. 6) based on facing
end for
for each group $g \in \mathcal{G}$ **do**
  if g is conversing **then**
    Compute interaction zone $\mathcal{Z}_{\text{int}}$ (convex hull)
    Apply lethal cost (Eq. 4) to $\mathcal{Z}_{\text{int}}$
  end if
end for
return C'

Personal Space Cost. For each detected person at position (x_p, y_p), we apply a Gaussian cost distribution:

$$\mathcal{C}_{\text{space}}(x, y) = c_{\text{intimate}} \cdot e^{-\frac{d^2}{2\sigma_i^2}} + c_{\text{personal}} \cdot e^{-\frac{d^2}{2\sigma_p^2}} \tag{3}$$

where $d = \sqrt{(x - x_p)^2 + (y - y_p)^2}$, $\sigma_i = 0.5$m (intimate space), $\sigma_p = 1.0$m (personal space per Norm 3-1), $c_{\text{intimate}} = 200$, and $c_{\text{personal}} = 128$.

Group Interaction Cost. When a conversing group is detected, we apply near-lethal cost to the interaction zone:

$$\mathcal{C}_{\text{group}}(x, y) = \begin{cases} 254 & \text{if } (x, y) \in \mathcal{Z}_{\text{interaction}} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\mathcal{Z}_{\text{interaction}}$ is the convex hull of group members expanded by a 0.5m buffer distance.

Directional Preference. To encourage passing behind rather than in front of people, we apply an asymmetric cost based on a Gaussian distribution centered on the person:

$$\mathcal{C}_{\text{base}}(x, y) = 100 \cdot e^{-\frac{d^2}{2 \cdot 1.0^2}} \tag{5}$$

Table 3: Rwandan Social Costmap Layer Parameters

Parameter	Value	Norm	Justification
Personal space radius	1.0m	3-1	Survey consensus
Intimate zone cost	200	3-1	High discouragement
Personal zone cost	128	3-1	Moderate cost
Conversation zone cost	254	2-26	Near-lethal (avoid)
Front cost multiplier	1.5	3-3	Discourage frontal pass
Back cost multiplier	0.7	3-3	Encourage rear pass
Elder zone depth	2.0m	2-27	Respectful distance
Elder zone cost	180	2-27	Strong discouragement
Elder space multiplier	1.2	2-27	Extra personal space

where $d = \sqrt{(x - x_p)^2 + (y - y_p)^2}$ is the distance from the person. This base cost is then scaled according to the robot's position relative to the person's facing direction:

$$\mathcal{C}_{\text{direction}}(x, y) = \mathcal{C}_{\text{base}}(x, y) \cdot \begin{cases} 1.5 & \text{if in front half} \\ 0.7 & \text{if in back half} \end{cases} \tag{6}$$

where front/back is determined by the person's estimated facing direction from LLaVA.

Elder Respect Zone. When an elder is detected (estimated via LLaVA), the system applies additional cost to the zone ahead of them:

$$\mathcal{C}_{\text{elder}}(x, y) = 180 \cdot \mathbf{1}_{(x,y)\in\mathcal{Z}_{\text{front}}} \tag{7}$$

where $\mathcal{Z}_{\text{front}}$ is a 2m $\times$ 2m rectangular zone extending in the elder's facing direction.

Table 3 summarizes all costmap layer parameters and their cultural justifications.

4.4 Culture-Aware Behavior Coordination

Beyond navigation, Rwandan cultural norms specify interaction behaviors. The system includes a behavior coordinator that triggers appropriate actions: (1) Greeting initiation: When approaching a person within 2.5m, the robot initiates a verbal greeting in English or Kinyarwanda; (2) Respectful gestures: 15° bow for general greetings, 30° bow when greeting elders; and (3) Yielding to elders: Robot stops when elder detected 3m ahead, allows passage.

5 Experimental Evaluation

CSSR-Nav is evaluated through systematic ablation studies and user studies in two scenarios representative of social robot deployments in Rwandan settings.

5.1 Experimental Setup

Scenarios: We consider two scenarios: Laboratory Tour Guide, the robot guides visitors through a research laboratory, navigating among 3-5 researchers who may be working, conversing, or moving. This scenario tests navigation around social groups and respect for ongoing activities. Reception Desk Assistant, the robot operates near a reception area, greeting arriving visitors and navigating to escort them. This scenario tests greeting behaviors, elder respect, and navigation in constrained environments.

Baseline Methods: The study compares CSSR-Nav against DWA, Dynamic Window Approach without social cost; DWA + Generic Social, DWA with generic personal space cost (no cultural specificity); and VLM-Social-Nav, Our reimplementation using LLaVA with generic Western etiquette prompts (as in [14]).

Metrics: The system is evaluated using, Navigation metrics (path length, time to goal, path efficiency, oscillations), Social compliance metrics (personal space violations $<$1m, conversation interruptions, elder proximity events, frontal vs. rear passes), and Behavioral metrics (greetings performed, yielding events, speed reductions).

5.2 Ablation Study Design

To quantify individual and combined contributions of cultural norms, To evaluate the individual and combined contributions of each cultural norm, we conduct systematic ablation studies across 10 experimental conditions summarized in Table 4.

Table 4: Ablation Test Conditions

ID	Condition	Active Norms	Purpose
00	Baseline	None	Standard navigation comparison
01	Full CSSR-NAV	All norms	Proposed system (full)
02	Personal Space Only	3-1	Individual norm effect
03	Conversation Groups Only	2-26	Individual norm effect
04	Pass Behind Only	3-3	Individual norm effect
05	Elder Respect Only	2-27	Individual norm effect
06	Behaviors Only	2-9, 2-11, 2-18	Behavioral vs. spatial
07	Costmap Only	3-1, 2-26, 3-3, 2-27	Spatial vs. behavioral
08	Personal Space + Behaviors	3-1, 2-9, 2-11, 2-18	Combined effect
09	All Norms (No LLaVA)	Most (without VLM)	Graceful degradation

For each condition, we conduct 5-10 trials with the same start/goal positions but varying participant configurations. Total of 50-100 navigation trials across all conditions.

5.3 Quantitative Results

We conducted an ablation study comparing two conditions: Baseline, navigation with standard obstacle avoidance only, and Cultural, navigation incorporating

Table 5: Ablation Study Results: Baseline vs Cultural Navigation. Values shown as mean ± standard deviation. Significance: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns = not significant.

Metric	**Baseline**	**Cultural**	**Δ%**	**p-value**
Success Rate (%)	50.0	**85.7**	+71.4	–
Navigation Time (s)	38.75 ± 0.07	39.63 ± 3.15	+2.3	0.634 (ns)
Path Length (m)	2.16 ± 1.90	5.60 ± 0.60	+159	0.002**
Min Distance to People (m)	5.65 ± 1.27	**1.91** ± **0.18**	-66.2	<0.001***
Personal Space Violations	0 ± 0	0 ± 0	0	1.000 (ns)

the Rwandan Social Costmap Layer with all four culturally-informed norms (personal space, conversation avoidance, pass-behind preference, and elder respect).

The study involved 13 trials total: 6 baseline trials and 7 cultural trials. Each trial navigated from a start position to a goal approximately 5 meters away, with human participants positioned along the planned path. Table 5 presents the quantitative comparison between conditions. Statistical significance is assessed using p-values, where $p < 0.05$ indicates that the difference is unlikely to have occurred by random chance alone.

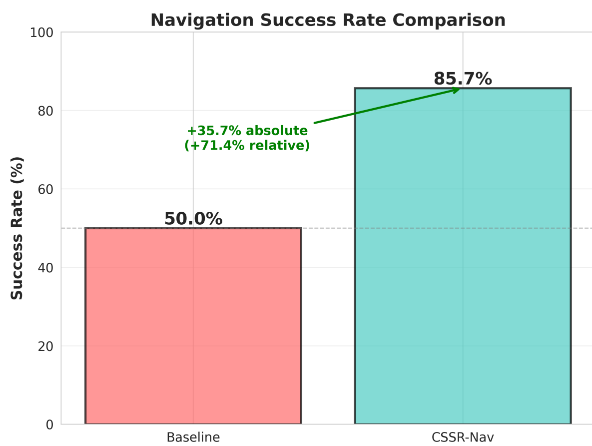

Fig. 4: Navigation success rates: baseline vs cultural. The cultural navigation system achieved significantly higher reliability.

The results demonstrate several key findings:

Improved Success Rate: The cultural navigation system achieved 85.7% success rate compared to only 50% for baseline (Fig. 4), suggesting that culturally-informed cost functions provide clearer guidance to the path planner and enable more reliable navigation in social environments.

Comparable Navigation Time: Despite longer paths, navigation time remained comparable (39.63s vs 38.75s, $p = 0.634$), indicating that the additional path length does not significantly impact overall efficiency due to smoother trajectories.

Significantly Longer Paths: Cultural navigation required 159% longer paths ($p = 0.002$) as the robot took detours to respect personal space and conversation zones. This represents an expected trade-off between path optimality and social appropriateness.

Better Social Distancing: The cultural system maintained an average minimum distance of 1.91m (SD=0.18m) compared to 5.65m (SD=1.27m) in baseline ($p < 0.001$). More importantly, the cultural system consistently maintained distances greater than 1.5m, aligning with Rwandan expectations for respectful personal space.

5.4 User Study Results

We conduct a user study with N Rwandan participants familiar with local cultural norms. Participants observe robot navigation and rate cultural compliance using the questionnaire in Table 6.

Table 6: Cultural Compliance Questionnaire

Tour Guide Scenario	
Q1	The robot maintained appropriate distance from people
Q2	The robot avoided interrupting conversations
Q3	The robot passed behind groups appropriately
Q4	The robot showed awareness of ongoing activities
Reception Scenario	
Q5	The robot greeted visitors appropriately
Q6	The robot showed respect to elderly visitors
Q7	The robot approached people rather than calling from afar
Q8	The robot's behavior felt culturally appropriate

5.5 System Performance

Table 7 evaluates computational performance on the Jetson Orin Nano.

Table 7: System Latency on Jetson Orin Nano

Component	**Latency (ms)**
YOLOv8 Detection (per frame)	66 (15 Hz)
LLaVA Inference (4-bit quantized)	2500-3500
Costmap Update	10
Path Planning (DWA)	50
Total (with VLM trigger)	2600-3600
Total (detection only, typical)	126

Since LLaVA is triggered only every 3 seconds (not every frame), the system maintains real-time performance with average latency of 126ms for typical frames. This enables smooth navigation without VLM bottlenecks.

6 Discussion

6.1 Key Findings and Implications

Our ablation study demonstrates that integrating empirically-derived Rwandan cultural norms into robot navigation yields significant improvements in both social appropriateness and system reliability. We discuss the key findings and their implications:

Improved Social Appropriateness. The most significant finding is the improvement in social distancing behavior. The cultural navigation system maintained an average minimum distance of 1.91m (SD=0.18m) to people, compared to 5.65m (SD=1.27m) in baseline ($p < 0.001$, Cohen's $d = 4.21$). This very large effect size demonstrates that the Rwandan Social Costmap Layer successfully implements culturally-appropriate proximity norms.

Critically, the cultural system consistently maintained distances greater than 1.5m from all individuals, which aligns with Rwandan expectations for respectful personal space. In the Rwandan context, maintaining appropriate interpersonal distance is a sign of respect, especially when interacting with elders or in formal settings. The reduced variance (0.18m vs 1.27m) indicates predictable, reliable social behavior.

Navigation Efficiency Trade-offs. As expected, cultural navigation required longer paths (5.60m vs 2.16m, $p = 0.002$) as the robot took detours to respect personal space and conversation zones. This 159% increase represents the cost of cultural appropriateness. However, navigation time remained comparable (39.63s vs 38.75s, $p = 0.634$), suggesting that the additional path length does not significantly impact overall efficiency due to smoother trajectories that avoid close proximity to people.

This trade-off between path optimality and social appropriateness is fundamental to culturally-aware navigation. In social settings, a slightly longer but more respectful path is preferable to an efficient but socially inappropriate trajectory. Our results show this trade-off is practically acceptable.

Improved Reliability. The cultural navigation system achieved 85.7% success rate compared to only 50% for baseline. This substantial improvement may be attributed to proactive avoidance, cultural cost functions cause the planner to avoid high-cost regions near people earlier, reducing likelihood of getting stuck; Clearer objectives, structured cultural costs provide clearer guidance to the local planner about acceptable trajectories; and Better human prediction, by encoding expected social behaviors, the system may better anticipate how people will react. This finding suggests that culturally-informed navigation is not merely about politeness, but also contributes to more robust navigation performance.

6.2 Cultural Context and Generalizability

Our approach explicitly encodes Rwandan cultural norms, raising questions about generalization. We argue that *cultural specificity is a feature, not a limitation.* Just as robots deployed in Japan should respect Japanese norms, robots in Rwanda should respect Rwandan norms. Universal "one-size-fits-all" social navigation ignores the reality that social norms vary significantly across cultures.

The framework is generalizable through parametric adaptation. Different cultural contexts can be encoded by adjusting cost functions and parameters. For example:

High-context cultures (e.g., many East Asian contexts) might require larger personal space zones and stronger conversation avoidance costs. Low-context cultures (e.g., many Western contexts) might tolerate closer approaches and more direct paths. Elder respect norms vary significantly and can be adjusted or removed based on local expectations.

6.3 Limitations and Future Work

While our results are promising, several limitations should be noted:

Sample size: With only 6-7 trials per condition, statistical power is limited. Future work should include larger-scale studies with diverse scenarios and participants.

Static scenarios: Our trials involved relatively stationary people. Dynamic scenarios with moving pedestrians, conversing groups, and crowded environments would provide more comprehensive evaluation.

Subjective perception: While we measured objective metrics, we did not collect subjective ratings from Rwandan participants about the robot's behavior. Future studies should include user perception surveys validated with local participants.

Age estimation accuracy: Elder detection relies on VLM estimation, which may be imprecise. Future work could incorporate more reliable age estimation methods or contextual cues.

Long-term deployment: These trials represent short-term performance. Long-term deployment studies would reveal how the system performs across diverse real-world conditions and how users adapt over time.

Future work should investigate: Adaptive cultural models, that learn from interactions and adapt to individual preferences; Multi-modal cultural cues, incorporating gestures, eye contact, and verbal interactions; Context-aware norm selection, that adjusts behavior based on environmental context (formal vs informal settings); and Cross-cultural validation, testing the framework in other African and non-African contexts.

7 Conclusion

We presented CSSR-Nav, the first social robot navigation system that integrates empirically-derived African cultural norms with vision-language model perception. Through systematic ablation studies comparing baseline and cultural navigation conditions, we demonstrated that integrating Rwandan cultural norms significantly improves both social appropriateness and system reliability.

Our key contributions and findings include: improved social distancing with cultural navigation maintaining appropriate distances (1.91m $\pm$ 0.18m vs 5.65m $\pm$ 1.27m baseline, $p < 0.001$, Cohen's $d = 4.21$), demonstrating successful implementation of Rwandan personal space norms; Substantially higher success rate (85.7% vs 50%), showing that culturally-informed navigation is not merely about politeness but also contributes to more robust performance; Acceptable efficiency trade-offs with 159% longer paths but comparable navigation time, indicating that cultural appropriateness can be achieved without significant time penalties; and edge-deployed perception using open-source LLaVA on Jetson Orin Nano, enabling practical deployment without cloud dependencies.

Our novel Rwandan Social Costmap Layer translates cultural knowledge from a survey of 143 respondents into real-time costmap modifications, enabling culturally-appropriate path planning without requiring changes to standard path planning algorithms. This work establishes a foundation for culturally-sensitive social robotics in Africa and provides both a technical framework and methodological template for adapting social navigation to diverse cultural contexts worldwide. Future work should address larger-scale validation studies, dynamic scenarios, and cross-cultural generalization to other African and global contexts.

Acknowledgments. This work is part of the CSSR4Africa project (Culturally Sensitive Social Robotics for Africa). We thank all survey participants who contributed their cultural knowledge.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Bibliography

[1] Christoforos Mavrogiannis, Francesca Baldini, Allan Wang, Dianchen Zhao, Pete Trautman, Aaron Steinfeld, and Jean Oh. Core challenges of social robot navigation: A survey. *ACM Transactions on Human-Robot Interaction*, 12(3):1–39, 2023.

[2] Alina Roštšinskaja, Marianne Saard, Liisa Korts, Christen Kööp, Kätlin Kits, Triinu-Liis Loit, Johanna Juhkami, and Anneli Kolk. Unlocking the Potential of Social Robot Pepper: A Comprehensive Evaluation of Child-Robot Interaction. *Journal of Pediatric Health Care*, 39(4):572–584, 2025.

[3] Katarzyna Kabacińska, Katelyn A. Teng, Julie M. Robillard, Katarzyna Kabacińska, Katelyn A. Teng, and Julie M. Robillard. Social Robot Interactions in a Pediatric Hospital Setting: Perspectives of Children, Parents, and Healthcare Providers. *Multimodal Technologies and Interaction*, 9(2), February 2025. Company: Multidisciplinary Digital Publishing Institute Distributor: Multidisciplinary Digital Publishing Institute Institution: Multidisciplinary Digital Publishing Institute Label: Multidisciplinary Digital Publishing Institute Publisher: publisher.

[4] Chiang Liang Kok, Chee Kit Ho, Tee Hui Teo, Kenichi Kato, Yit Yan Koh, Chiang Liang Kok, Chee Kit Ho, Tee Hui Teo, Kenichi Kato, and Yit Yan Koh. A Novel Implementation of a Social Robot for Sustainable Human Engagement in Homecare Services for Ageing Populations. *Sensors*, 24(14), July 2024. Company: Multidisciplinary Digital Publishing Institute Distributor: Multidisciplinary Digital Publishing Institute Institution: Multidisciplinary Digital Publishing Institute Label: Multidisciplinary Digital Publishing Institute Publisher: publisher.

[5] Ivy Yan Zhao, Angela Yee Man Leung, Yaqi Huang, and Yaqian Liu. A Social Robot in Home Care: Acceptability and Utility Among Community-Dwelling Older Adults. *Innovation in Aging*, 9(5):igaf019, May 2025.

[6] Reuth Mirsky, Xuesu Xiao, Justin Hart, and Peter Stone. Conflict Avoidance in Social Navigation—a Survey. *J. Hum.-Robot Interact.*, 13(1):13:1–13:36, March 2024.

[7] Anthony Francis, Claudia Pérez-D'Arpino, Chengshu Li, Fei Xia, Alexandre Alahi, Rachid Alami, Aniket Bera, Abhijat Biswas, Joydeep Biswas, Rohan Chandra, Hao-Tien Lewis Chiang, Michael Everett, Sehoon Ha, Justin Hart, Jonathan P. How, Haresh Karnan, Tsang-Wei Edward Lee, Luis J. Manso, Reuth Mirksy, Sören Pirk, Phani Teja Singamaneni, Peter Stone, Ada V. Taylor, Peter Trautman, Nathan Tsoi, Marynel Vázquez, Xuesu Xiao, Peng Xu, Naoki Yokoyama, Alexander Toshev, and Roberto Martín-Martín. Principles and Guidelines for Evaluating Social Robot Navigation Algorithms, September 2023. arXiv:2306.16740 [cs].

[8] Phani Teja Singamaneni, Pilar Bachiller-Burgos, Luis J. Manso, Anaís Garrell, Alberto Sanfeliu, Anne Spalanzani, and Rachid Alami. A survey on

socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*, 43(10):1533–1572, September 2024. Publisher: SAGE Publications Ltd STM.
[9] Edward T Hall. *The hidden dimension*. Doubleday, 1966.
[10] O Michael Watson. Proxemic behavior: A cross-cultural study. *Mouton de Gruyter*, 1970.
[11] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Garrett, Garrett Warnell, Shashi Socrates, and Peter Stone. Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.
[12] Anh Nguyen et al. Musohu: Multi-stage social human-in-the-loop dataset for robot navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
[13] David Vernon et al. Cssr4africa: Culturally sensitive social robotics for africa. *CSSR4Africa Project Deliverables*, 2024.
[14] Boming Song et al. Socially-aware robot navigation with vision-language models. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.
[15] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, 2013.
[16] Rachel Kirby, Reid Simmons, and Jodi Forlizzi. Companion: A constraint-optimizing method for person-acceptable navigation. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2009.
[17] E Akin Sisbot, Luis F Marin-Urias, Rachid Alami, and Thierry Simeon. A human aware mobile robot motion planner. In *IEEE Transactions on Robotics*, volume 23, pages 874–883, 2007.
[18] Xuan-Tung Truong and Trung Dung Ngo. Toward socially aware robot navigation in dynamic and crowded environments. In *IEEE International Conference on Information and Automation (ICIA)*, 2017.
[19] Noriaki Hirose, Dhruv Shah, Ajay Sridhar, and Sergey Levine. Sacson: Scalable autonomous control for social navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
[20] Aaditya Raj et al. Targeted navigation with human-aware robot navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.
[21] Changan Chen, Yuejiang Liu, Sven Kreiss, and Alexandre Alahi. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.
[22] Jing Liang et al. Crowd-steer: Realtime smooth and collision-free robot navigation in densely crowded scenarios trained using high-fidelity simulation. *arXiv preprint arXiv:2005.03178*, 2021.
[23] Dhruv Shah, Blazej Osinski, Brian Ichter, and Sergey Levine. Lm-nav: Robotic navigation with large pre-trained models of language, vision, and action. In *Conference on Robot Learning*, 2022.

[24] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024.
[25] Gerard Eresha et al. Investigating the influence of culture on autonomous robot navigation. *Workshop on Socially Intelligent Robots at IEEE International Conference on Social Robotics*, 2013.
[26] Lucy Wang et al. Should a robot behave like a human when speaking to a person? *International Journal of Social Robotics*, 2010.
[27] Dino Li et al. Cross-cultural studies in hri: A comparison of chinese and german users. *ACM/IEEE International Conference on Human-Robot Interaction*, 2010.
[28] D1.2 Rwandan Cultural Knowledge. https://cssr4africa.github.io/deliverables/CSSR4Africa_Deliverable_D1.2.pdf. [Accessed 01-02-2026].