

Ministère de l'enseignement Supérieur et de la recherche Scientifique

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculté de Technologie



جامعة باجي مختار – عنابة

كلية التكنولوجيا

Département Electronique

قسم الالكترونيات

Thèse

Présentée pour obtenir le diplôme de

Doctorat 3^{ème} cycle

Département : Electronique

Filière : Electronique

Spécialité : Électronique Médicale

Par :

BOUGHELOUM LOUBNA

Thème :

Un système d'assistance à la mobilité des non-voyants.

Thèse soutenue le date de soutenance devant le jury composé de :

N°	Nom et prénom	Grade	Etablissement	Qualité
01	Pr. FEZARI Mohamed	Prof.	Université Badji Mokhtar -Annaba	Président
02	Pr. BOUSBIA-SALAH Mounir	Prof.	Université Badji Mokhtar -Annaba	Rapporteur
03	Pr. BETTAYEB Mohamed	Prof.	Université Badji Mokhtar -Annaba	Co-rapporteur
04	Pr. AMARA KORBA Mohamed Cherif	Prof.	Professeur à l'Université Mohamed Cherif Messaadia-Souk Ahras	Examineur
05	Pr. BOUROUBA Hocine	Prof.	Université 8 mai 1945-Guelma	Examineur
06	Dr. BENSAOULA SALAH	MCA	Université Badji Mokhtar -Annaba	Examineur

2025/2026

Dédicaces

À mes parents, pour leur amour, leur patience et tous les sacrifices faits tout au long de mon parcours. Leur soutien constant m'a permis d'avancer et de surmonter les moments les plus difficiles.

À mes frères et sœurs, pour leur présence, leurs encouragements et leur confiance sincère.

À toutes celles et ceux qui m'ont accompagnée de près ou de loin, par un mot, une prière ou un geste, et qui ont contribué à faire de ce travail bien plus qu'un simple aboutissement académique.

Cette thèse est le reflet d'un long parcours fait de persévérance, de sacrifices et d'espoir.

Je vous la dédie avec une profonde gratitude et beaucoup d'émotion.

Loubna Bougheloum

Remerciements

Je remercie Dieu, qui m'a donné la force, la patience et le courage pour mener à bien ce travail, et m'avoir guidée à chaque étape sur le chemin de la connaissance.

Je tiens à exprimer ma profonde gratitude à mon encadrant, le Professeur « Bousbia-Salah Mounir », pour son engagement constant, ses conseils avisés et sa rigueur scientifique. Sa confiance et son accompagnement ont été essentiels pour surmonter les défis et atteindre les objectifs de cette thèse.

Je remercie également mon Co-encadrant, Professeur « Bettayeb Maamar », qui, malgré la distance, a toujours été présent et m'a apporté ses conseils précieux, sa disponibilité et ses encouragements motivants.

Je remercie les membres du jury pour avoir évalué ce travail.

Je remercie profondément ma famille pour son amour, sa patience et son soutien indéfectible, qui ont été ma motivation dans les moments difficiles.

Je remercie mes amis pour leur écoute, leur amitié sincère et leurs encouragements, qui ont rendu ce parcours plus agréable et porteur d'énergie.

Loubna Bougheloum

Résumé

La déficience visuelle limite la perception spatiale, l'anticipation des obstacles et la capacité à se déplacer de manière autonome, augmentant le risque de collisions et réduisant l'accès aux environnements complexes. Ces limitations nécessitent des systèmes intelligents capables de fournir des informations fiables sur l'environnement et des recommandations de trajectoire adaptées aux utilisateurs déficients visuels.

Un système modulaire et intégré a été développé, combinant détection, classification, localisation et évitement d'obstacles, avec un traitement en temps réel sur plateforme embarquée et des interfaces adaptées aux capacités sensorielles des utilisateurs. Le système indoor comprend la classification de pièces, la détection d'objets et la localisation relative, tandis que le système outdoor simule quatre sous-systèmes : détection, classification des obstacles statiques ou dynamiques, évitement et localisation.

Les résultats montrent que YOLOv7-Plus atteint une mAP de 0,87 sur la base WOTR, la classification par DINOv2+ConvLSTM atteint une précision de $91,7 \pm 0,9\%$, et le contrôleur flou simule des trajectoires anticipant tous les obstacles, en élargissant la trajectoire pour les obstacles dynamiques par rapport à une canne blanche. La fusion EKF+IMU+GPS assure une localisation précise avec un RMSE de $1,30 \times 10^{-6}$ m en espace structuré, 0,014 m en espace non structuré et 0,0011 m sur campus universitaire. Pour l'indoor, YOLOv8s maintient une mAP de 72,6 %, MobileNetV2 atteint 79,3 % de précision sur la classification des pièces, et la localisation relative présente une erreur MAE moyenne de 0,035 m. L'évaluation auprès de participants déficients visuels confirme la faisabilité opérationnelle et met en évidence des pistes d'amélioration pour l'interaction, le contrôle utilisateur et l'adaptation à des environnements complexes tels que centres commerciaux ou bibliothèques.

Ces travaux établissent un cadre méthodologique robuste pour la conception de systèmes d'assistance à la navigation, conciliant précision algorithmique, contraintes matérielles et besoins utilisateurs, et ouvrent la voie à une mobilité plus autonome et sécurisée dans des environnements indoor et outdoor.

Mots-clés : Déficience visuelle, système de navigation assistée, détection d'obstacles, classification des obstacles, évitement d'obstacle, localisation.

Abstract

Visual impairment restricts spatial perception, obstacle anticipation, and independent movement, increasing the risk of collisions and limiting access to complex environments. Addressing these challenges requires intelligent navigation systems that provide reliable environmental information and adaptive guidance for visually impaired users.

A modular and integrated navigation system was developed, combining obstacle detection, classification, localization, and avoidance, with real-time processing on an embedded platform and interfaces tailored to users' sensory capabilities. The indoor system integrates room classification, object detection, and relative localization, while the outdoor system simulates four subsystems: detection, classification of static or dynamic obstacles, avoidance, and localization.

Evaluation shows that YOLOv7-Plus achieves a mAP of 0.87 on the WOTR dataset, and DINOv2+ConvLSTM classification reaches an accuracy of $91.7 \pm 0.9\%$. The fuzzy logic controller anticipates all obstacles and generates wider trajectories for dynamic obstacles compared to a conventional white cane. EKF+IMU+GPS sensor fusion provides precise localization with RMSE values of 1.30×10^{-6} m in structured spaces, 0.014 m in unstructured spaces, and 0.0011 m on a university campus. For indoor navigation, YOLOv8s achieves a mAP of 72.6 %, MobileNetV2 reaches 79.3 % accuracy for room classification, and relative localization presents an average MAE of 0.035 m. User evaluations with visually impaired participants confirm operational feasibility and suggest improvements for interaction, user control, and adaptation to complex environments such as shopping malls, retail stores, and libraries.

These findings establish a robust framework for the design of assistive navigation systems that balance algorithmic precision, embedded hardware constraints, and user needs, supporting safer and more autonomous navigation in both indoor and outdoor environments.

Keywords : Visual impairment, assistive navigation system, obstacle detection, obstacle classification, obstacle avoidance, localization.

ملخص

تفيد الإعاقة البصرية الإدراك المكاني والقدرة على توقع العوائق والحركة المستقلة، مما يزيد من مخاطر الاصطدام ويقلل من إمكانية الوصول إلى البيانات المعقدة. ولعلاج هذه الصعوبات، تبرز الحاجة إلى أنظمة توجيه ذكية توفر معلومات بيئية موثوقة وإرشادات مرنة تتوافق مع قدرات المستخدمين البصرية.

تم تطوير نظام توجيه متكامل ومجزأ، يجمع بين كشف العوائق وتصنيفها وتحديد المواقع وتجنب التصادم، مع معالجة البيانات في الوقت الفعلي على منصة مدمجة وواجهات مصممة خصيصاً للحواس المتاحة للمستخدم. يدمج النظام الداخلي تصنيف الغرف وكشف الأشياء وتحديد الموقع النسبي، بينما يحاكي النظام الخارجي أربعة أنظمة فرعية: الكشف، وتصنيف العوائق الثابتة أو المتحركة، والتجنب، وتحديد الموقع.

أظهرت نتائج التقييم أن نموذج YOLOv7-Plus حقق دقة متوسطة (mAP) بلغت 0.87 على قاعدة بيانات WOTR، بينما وصلت دقة التصنيف باستخدام DINOv2+ConvLSTM إلى $91.7 \pm 0.9\%$. كما أثبت نظام التحكم بالمنطق الضبابي قدرته على التنبؤ بكافة العوائق ورسم مسارات أكثر اتساعاً وتجنباً للعوائق المتحركة مقارنة بالعصا البيضاء التقليدية.

أما عملية دمج المستشعرات (EKF+IMU+GPS)، فقد وفرت تحديداً دقيقاً للموقع بقيم خطأ تربيعي متوسط (RMSE) بلغت 1.30×10^{-6} م في المساحات المنظمة، و 0.014 م في المساحات غير المنظمة، و 0.0011 م داخل الحرم الجامعي. وفيما يخص التحرك الداخلي، حقق نموذج YOLOv8s دقة متوسطة بلغت 72.6%، ووصلت دقة نموذج MobileNetV2 في تصنيف الغرف إلى 79.3%، مع متوسط خطأ مطلق (MAE) في تحديد الموقع النسبي قدره 0.035 م.

أكدت تقييمات المستخدمين من ذوي الإعاقة البصرية الجدوى التشغيلية للنظام، واقترحت تحسينات في التفاعل والتحكم والتكيف مع البيانات المعقدة مثل مراكز التسوق والمكتبات. تؤسس هذه النتائج لإطار عمل قوي لتصميم أنظمة توجيه مساعدة توازن بين دقة الخوارزميات، وقيود الأجهزة، واحتياجات المستخدم، مما يدعم التحرك الآمن والمستقل في البيئات الداخلية والخارجية.

الكلمات المفتاحية: الإعاقة البصرية، أنظمة التوجيه المساعدة، كشف العوائق، تصنيف العوائق، تجنب العوائق، تحديد الموقع.

Table des matières

Résumé	i
Table des matières	iv
Liste des Tableaux	ix
Liste des Figures	x
Nomenclatures	xii
Introduction Générale	1
 Chapitre 1 : État de l'Art des Systèmes d'Assistance à la Mobilité pour Personnes Déficiences Visuelles	5
1.1. Introduction	6
1.2. La Mobilité des Personnes Déficiences Visuelles	6
1.2.1. Cognition spatiale sans vision	6
1.2.1.1. Intégration sensorielle non visuelle et traitement spatial.....	7
1.2.1.2. Mécanismes de mémorisation spatiale sans vision.....	8
1.2.1.3. Principes cognitifs directeurs pour la conception des systèmes d'assistance à la mobilité	9
1.2.1.4. Exploitation cognitive des représentations spatiales chez les personnes non-voyantes	10
1.2.2. Gestion de la navigation selon les degrés de déficience visuelle	11
1.2.3. Défis et difficultés de la vie quotidienne des personnes non-voyantes	13
1.3. Aides traditionnelles et leurs limites	15
1.3.1. La canne blanche	15
1.3.2. Le chien guide l'intelligence biologique	16
1.3.3. Insuffisances fonctionnelles des aides traditionnelles et émergence de nouveaux besoins d'assistance	17
1.4. Évolution des Systèmes d'Assistance Technologiques Des Capteurs à l'IA.....	19
1.4.1. Première Génération : L'ère des Sens Électroniques Élémentaires	19
1.4.1.1. L'Écholocalisation Ultrasonore	19
1.4.1.2. La triangulation Optique Infrarouge (Mesurer par la Lumière)	19
1.4.1.3. Laser par Temps de Vol	20
1.4.2. Deuxième Génération (1995–2010) : Navigation Géoréférencée et Cartographie Numérique	20
1.4.2.1. Le GPS et Trilatération Temporelle	20
1.4.2.2. PDA et des Systèmes Intégrés	21
1.4.2.3. La Transition Vers le Smartphone et les Applications Modernes	21

1.4.3. Troisième Génération (2010-Présent) Perception Intelligente et de la Compréhension Sémantique.....	21
1.4.3.1. Émergence des Systèmes de Vision Assistée	21
1.4.3.2. Intégration de l'apprentissage profond et optimisation systémique des architectures.....	22
1.4.3.3. Intégration Systémique, Fusion Multi-Capteurs et IA Efficiente	22
1.5. Critères d'évaluation des systèmes d'assistance à la mobilité	23
1.5.1. Performance fonctionnelle globale	23
1.5.2. Précision, fiabilité et cohérence temporelle	24
1.5.3. Réactivité et latence.....	24
1.5.4. Charge cognitive et clarté informationnelle.....	25
1.5.5. Sécurité et maîtrise des situations critiques	25
1.5.6. Acceptabilité et confort d'usage	25
1.5.7. Adaptabilité et capacité de généralisation.....	26
1.5.8. Cadres d'évaluation et limites méthodologiques	26
1.5.9. Enjeux actuels de l'évaluation des systèmes d'assistance.....	27
1.6. Conclusion.....	27
 Chapitre 2 : Fondements théoriques et modélisation du système d'assistance à la mobilité	28
2.1. Introduction	29
2.2. Définition d'un système d'assistance à la mobilité	29
2.3. Représentation de l'environnement pour la mobilité non-visuelle.....	30
2.3.1. Composants essentiels pour la modélisation de l'environnement non-visuel	30
2.3.2. Typologie des environnements et composantes spatiales	31
2.3.3. Propriétés critiques des obstacles et dynamique environnementale	33
2.4. Besoins des personnes non voyantes dans un système d'assistance à la mobilité.....	34
2.4.1. Sécurité du déplacement et prévention des situations à risque.....	34
2.4.2. Orientation spatiale et compréhension progressive de l'environnement	35
2.4.3. Gestion de la charge cognitive et contrôle de l'information	36
2.4.4. Confiance, acceptabilité et continuité d'usage.....	37
2.5. Modélisation fonctionnelle du système	38
2.5.1. Perception de l'environnement	38
2.5.2. Localisation de l'utilisateur	40
2.5.3. Évitement d'obstacles.....	43
2.5.4. Restitution de l'information.....	44
2.5.5. Interactions et flux d'information	46
2.6. Principes ergonomiques et cognitifs pour l'assistance à la mobilité	47
2.7. Cadre conceptuel de conception de la plateforme d'assistance à la mobilité	49

2.7.1. Contraintes biomécaniques et intégration corporelle du dispositif.....	49
2.7.2. Architecture perceptive embarquée et positionnement des capteurs.....	50
2.7.3. Exigences computationnelles et garanties de fonctionnement en temps réel	51
2.7.4. Impact cognitif, apprentissage utilisateur et appropriation fonctionnelle.....	52
2.7.5. Rôle de la plateforme d'implémentation dans l'architecture globale du système	52
2.8. Conclusion	52

Chapitre 3 : Méthodes d'optimisation et stratégies algorithmiques pour les sous-systèmes d'assistance à la mobilité

3.1. Introduction	55
3.2. Architecture générale du système	56
3.3. Méthodologie du sous-système de détection.....	58
3.3.1. Détection d'obstacles en temps réel.....	58
3.3.2. Fonctionnement et architecture de YOLOv7	61
3.3.3. Optimisation de YOLOv7 pour la détection d'obstacles	63
3.3.3.1. Renforcement du réseau d'extraction de caractéristiques pour une extraction de caractéristiques enrichie et préservée	64
3.3.3.2. Optimisation du réseau d'agrégation multi-échelles pour une fusion multi-échelles améliorée et équilibrée	64
3.3.3.3. Entraînement du modèle	66
3.4. Méthodologie du sous-système de classification	67
3.4.1. Principe de fonctionnement de ConvLSTM pour la modélisation temporelle	67
3.4.2. Principe de fonctionnement de DINOv2 pour l'extraction de caractéristiques spatiales	69
3.4.3. Méthodologie hybride proposée	70
3.5. Méthodologie du sous-système d'évitement d'obstacles	73
3.5.1. Évitement des obstacles en temps réel.....	73
3.5.2. Principe général du contrôleur flou	74
3.5.3. Contrôleur flou proposé pour l'évitement en temps réel.....	75
3.5.3.1. Chaîne décisionnelle globale et cohérence inter-sous-systèmes.....	76
3.5.3.2. Variables d'entrée et de sortie du contrôleur flou	77
3.5.3.3. Base de règles floues	78
3.6. Méthodologie du sous-système de localisation.....	79
3.6.1. Principes théoriques du filtre de Kalman étendu pour l'estimation d'état.....	79
3.6.2. Approche de localisation proposée basée sur la fusion IMU–GPS	81
3.6.2.1. Prétraitement, calibrage et synchronisation des capteurs	82
3.6.2.2. Modélisation de l'état du système.....	82
3.6.2.3. Adaptation dynamique des covariances	83

3.7. Interface auditive et implémentation embarquée	84
3.7.1. Génération du retour audio contextuel pour l'assistance à la navigation	84
3.7.2. Choix et justification de la plateforme embarquée.....	85
3.8. Conclusion	85
 Chapitre 4 : Conception d'un système de navigation indoor pour les non-voyants	87
4.1. Introduction	88
4.2. Architecture globale du système.....	89
4.2.1. Principes architecturaux et structuration fonctionnelle du système	89
4.2.2. Architecture matérielle embarquée.....	90
4.2.3. Architecture logicielle et organisation des flux fonctionnels	91
4.3. Implémentation fonctionnelle des sous-systèmes	92
4.3.1. Classification contextuelle des environnements indoor	93
4.3.2. Détection des obstacles.....	95
4.3.3. Localisation et estimation du mouvement en environnement indoor	96
4.3.4. Génération d'indications vocales contextuelles	98
4.4. Conception et réalisation du système	99
4.5. Protocole expérimental et évaluation du système	101
4.6. Conclusion.....	103
 Chapitre 5 : Évaluation expérimentale et analyse des performances des systèmes de navigation assistée.....	105
5.1. Introduction	106
5.2. Évaluation du système de navigation outdoor simulé	106
5.2.1. Évaluation du sous-système de détection	107
5.2.2. Évaluation du sous-système de classification environnementale	108
5.2.3. Évaluation du sous-système d'évitement.....	110
5.2.4. Évaluation du sous-système de localisation outdoor.....	111
5.3. Évaluation du système de navigation indoor embarqué	113
5.3.1. Évaluation de la classification des pièces indoor	113
5.3.2. Évaluation de la détection d'objets indoor.....	114
5.3.3. Évaluation de la localisation et de l'évitement indoor	115
5.3.4. Intégration du système et faisabilité opérationnelle sur plateforme embarquée	116
5.3.5. Étude comparative avec des systèmes de navigation indoor existants	117
5.3.6. Etude utilisateur.....	118
5.4. Analyse globale des résultats et discussion.....	119
5.4.1. Résultats et discussion des performances du système outdoor	119

5.4.1.1. Résultat et discussion des performance du sous-système de détection.....	120
5.4.1.2. Résultats et discussion des performances du sous-système de classification	122
5.4.1.3. Résultats et discussion des performances du sous-système d'évitement.....	125
5.4.1.4. Résultats et discussion des performances du sous-système de localisation.....	127
5.4.2. Résultats et discussion des performances du système indoor	129
5.4.2.1. Résultats et discussion des performances du sous-système de classification des pièces	130
5.4.2.2. Résultats et discussion des performances du sous-système de détection d'objet	131
5.4.2.3. Résultats et discussion des performances du sous-système de localisation.....	132
5.4.2.4. Résultats et discussion des performances d'intégration système et faisabilité opérationnelle embarquée.....	134
5.4.2.5. Résultats et discussion de l'étude comparative avec des systèmes de navigation indoor existants	135
5.4.2.6. Résultat de l'étude utilisateur	137
5.5. Limites et perspectives d'amélioration.....	140
5.5.1. Limites du système outdoor.....	140
5.5.2. Limites du système indoor.....	141
5.5.3. Améliorations futures	142
5.6. Conclusion.....	143
Conclusion Générale	144
Références.....	147
Annexe A.....	154

Liste des Tableaux

Tableau 1. Comparaison entre méthodes de détection one-stage et two-stage	60
Tableau 2. Synthèse comparative des architectures YOLOv7 et YOLOv7-Plus.....	65
Tableau 3. Base de règles floues étendue avec gestion des situations critiques.....	78
Table 4. Résumé des profils des participants	101
Tableau 5. Comparaison globale des performances des modèles de détection	120
Tableau 6. Étude comparative des performances de classification.....	122
Tableau 7. Analyse détaillée des classifications par scène.....	123
Tableau 8. Étude comparative des performances de localisation (RMSE en mètres)	127
Tableau 9. Comparaison des modèles de classification de scènes indoor	130
Tableau 10. Résultats comparatifs des modèles de détection d'objets indoor.....	131
Tableau 11. Étude comparative centrée utilisateur avec les systèmes de navigation indoor existants.....	135
Tableau 12. Évaluation heuristique basée sur les principes ISO 9241-110	138

Liste des Figures

Figure 1. Anatomie du système visuel chez l'humain	2
Figure 1.1. La perception spatiale chez l'être humain	7
Figure 1.2. Les formes de malvoyances	12
Figure 1.3. Premier outil pour la navigation pour les non-voyants	16
Figure 1.4. Le TomPouce	20
Figure 2.1. Schéma général d'un système d'assistance à la mobilité pour les personnes non-voyantes	30
Figure 2.2. Typologie des environnements et composantes spatiales	32
Figure 2.3. Principe général de fonctionnement de la vision par ordinateur	39
Figure 2.4. Principe de fonctionnement des capteurs de distance	39
Figure 2.5. Principe de fonctionnement de la fusion multicapteurs	40
Figure 2.6. Le principe de la trilatération.....	41
Figure 2.7. Le modèle IMU pour la localisation	42
Figure 2.8. Un système général d'interférence d'un contrôleur flou	43
Figure 3.1. Architecture globale du système d'assistance à la mobilité pour personnes non voyantes	56
Figure 3.2. Principe de fonctionnement de YOLO	59
Figure 3.3. Architecture globale de YOLOv7	62
Figure 3.4. Architecture globale de la version optimisée	65
Figure 3.5. Architecture du ConvLSTM	68
Figure 3.6. Architecture générale de DINOv2	69
Figure 3.7. Architecture de la méthode hybride proposée pour la classification	71
Figure 3.8. Organigramme du contrôleur flou pour évitement d'obstacle.....	76
Figure 3.9. Fonctions d'appartenance des variables linguistiques d'entrées	77
Figure 3.10. Architecture de fusion IMU-GPS par Filtre de Kalman Etendu	81
Figure 4.1. Architecture générale du système de navigation indoor.....	89
Figure 4.2. Diagramme de l'architecture de classification indoor	94
Figure 4.3. Architecture du processus de détection d'obstacle pour environnement indoor	96
Figure 4.4. Diagramme du processus de localisation	97
Figure 4.5. Le prototype du système de navigation indoor	100
Figure 5.1. Résultat de performance de détection du modèle proposé	121
Figure 5.2. Résultat de la simulation du 1er Scénario d'évitement d'obstacle	125
Figure 5.3. Résultat de la simulation du 2eme Scénario d'évitement d'obstacle	125
Figure 5.4. Résultat de la simulation du 3eme Scénario d'évitement d'obstacle	126
Figure 5.5. Résultat de la simulation du 4eme Scénario d'évitement d'obstacle	126

Figure 5.6. Résultat de la simulation du sous-système de localisation	128
Figure 5.7. La matrice de confusion de la classification du modèle MobileNetv2	130
Figure 5.8. Résultats comparatifs du RMSE	133
Figure 5.9. Résultats comparatifs du MAE	133
Figure 5.10. Occupation des ressources et la latence des différents modules sur Raspberry Pi..	134
Figure 5.11. Le score du System Usability Scale (SUS)	137
Figure 5.12. Le score de du NASA Task Load Index (NASA-TLX).....	137
Figure 5.13. Bénévole qui utilise le système dans environnement indoor réelle	138

GPS : Global Positioning System (Système de positionnement mondial)

PGS : Personal Guidance System (Système de guidage personnel)

DGPS : GPS différentiel (Differential GPS)

GNSS : Global Navigation Satellite System (Système mondial de navigation par satellite)

IMU : Inertial Measurement Unit (Unité de mesure inertielle)

PDA : Personal Digital Assistant (Assistant numérique personnel)

RFID : Radio Frequency Identification (Identification par radiofréquence)

ToF : Time of Flight (Temps de vol)

EKF : Extended Kalman Filter (Filtre de Kalman étendu)

PF : Particle Filter (Filtre particulaire)

SLAM : Simultaneous Localization and Mapping (Localisation et cartographie simultanées)

LSD-SLAM : Large-Scale Direct Monocular SLAM (SLAM direct monoculaire à grande échelle)

CNN : Convolutional Neural Network (Réseau de neurones convolutif)

DINOv2 : DIstillation with NO labels version 2 (Distillation sans étiquettes version 2)

ViT : Vision Transformer (Transformeur visuel)

LSTM : Long Short-Term Memory (Mémoire à long et court terme)

ConvLSTM : Convolutional Long Short-Term Memory (LSTM convolutionnel)

RGB-D : Red Green Blue - Depth (Rouge Vert Bleu - Profondeur)

R-CNN : Region-based Convolutional Neural Network (Réseau de neurones convolutif basé sur les régions)

Fast R-CNN : Fast Region-based Convolutional Neural Network (Version rapide du R-CNN)

YOLO : You Only Look Once (Vous ne regardez qu'une fois)

YOLOv7-Plus : Version améliorée de YOLOv7

CBS : Convolution + Batch Normalization + SiLU (Convolution + Normalisation par lot + activation SiLU)

E-ELAN : Extended Efficient Layer Aggregation Network (Réseau d'agrégation de couches efficace étendu)

MP1 : Max Pooling layer 1 (Couche de pooling maximum 1)

PAFPN : Path Aggregation Feature Pyramid Network (Réseau pyramidal de caractéristiques avec agrégation de chemins)

SPPCSPC: Spatial Pyramid Pooling - Cross Stage Partial Channel (Pooling pyramidal spatial - canal partiel à étage croisé)

ELAN-H : Efficient Layer Aggregation Network - High (Réseau d'agrégation de couches efficace - version haute)

REP : Re-parameterization (Reparamétrisation)

IoU : Intersection over Union (Intersection sur union)

TP : True Positives (Vrais positifs)

FP : False Positives (Faux positifs)

FN : False Negatives (Faux négatifs)

mAP : mean Average Precision (Précision moyenne)

RMSE : Root Mean Square Error (Erreur quadratique moyenne)

MAE : Mean Absolute Error (Erreur absolue moyenne)

FPS : Frames Per Second (Images par seconde)

WOTR: Walk On The Road

SUS : System Usability Scale (Échelle d'utilisabilité du système)

ISO : International Organization for Standardization (Organisation internationale de normalisation)

TTS: Text-To-Speech (Synthèse vocale)

Introduction

Générale

Introduction Générale

La vision humaine représente l'un des systèmes sensoriels les plus élaborés du corps humain, tant par la sophistication de son organisation anatomique que par la complexité des mécanismes de traitement de l'information qu'elle mobilise. Comme illustré par la Figure 1, l'anatomie du système visuel repose sur une architecture intégrée comprenant l'œil, la rétine, et le nerf optique. La rétine assure la conversion de l'énergie lumineuse en signaux électriques grâce aux photorécepteurs, tandis que les cellules ganglionnaires transmettent ces signaux vers le cortex visuel primaire et les aires associatives. Ce traitement hiérarchique et distribué permet non seulement l'identification des formes, des couleurs et des mouvements, mais également l'interprétation sémantique et spatiale de la scène observée [1].

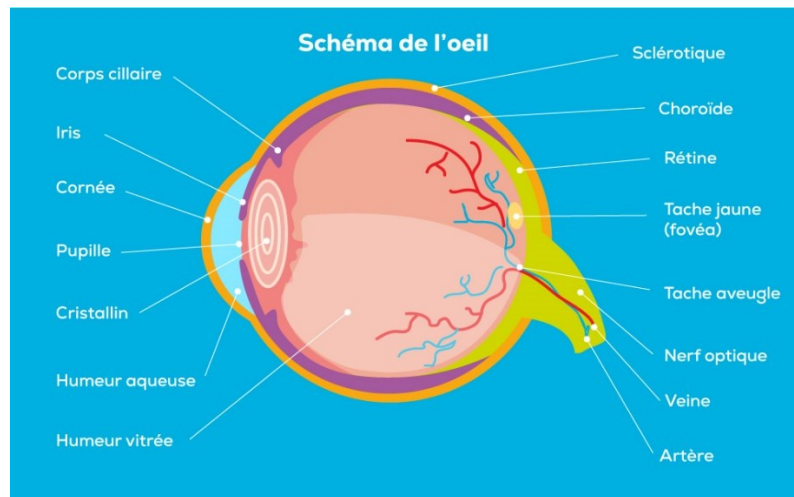


Figure 1. Anatomie du système visuel chez l'humain

La vision ne se limite donc pas à une fonction perceptive isolée ; elle constitue un pilier fondamental de la cognition spatiale et de l'interaction avec l'environnement. Elle participe à l'orientation, à l'évaluation des distances, à la détection et à l'anticipation des obstacles, ainsi qu'à la planification motrice. L'intégration dynamique entre vision périphérique, vision focale et mémoire visuelle assure une navigation fluide, intuitive et sécurisée dans des contextes variés [2].

Les déficiences visuelles, résultant de pathologies affectant l'une ou plusieurs composantes du système visuel représenté en Figure 1, perturbent profondément cet équilibre fonctionnel. Les atteintes rétiniennes peuvent altérer la vision centrale ou périphérique ; les neuropathies optiques compromettent la transmission des signaux visuels ; les lésions corticales peuvent affecter l'interprétation des informations reçues. Selon leur nature et leur sévérité, ces troubles entraînent une diminution de l'acuité visuelle, une réduction du champ visuel, une altération de la perception des contrastes ou une cécité complète [3].

Dans le cas de la non-voyance, la perte d'information visuelle modifie radicalement la relation à l'espace. Les mécanismes naturels d'anticipation, d'évitement et d'orientation doivent être compensés

par d'autres modalités sensorielles, principalement auditives et tactiles. Si la plasticité cérébrale permet une réorganisation fonctionnelle partielle, la navigation demeure une activité exigeante, sollicitant fortement l'attention, la mémoire spatiale et les capacités d'inférence [4].

Les environnements contemporains amplifient ces difficultés. Les espaces urbains extérieurs sont marqués par une densité élevée d'obstacles statiques et dynamiques, des variations d'éclairage importantes et une organisation spatiale parfois complexe. Les environnements intérieurs, tels que les bâtiments publics ou universitaires, présentent des configurations architecturales variées, des transitions fréquentes entre espaces et une absence de repères universels. Dans ces contextes, la mobilité autonome requiert une capacité constante d'identification des obstacles, d'estimation des distances et de compréhension de la structure spatiale environnante.

Les aides traditionnelles, notamment la canne blanche, offrent une détection fiable à courte portée, mais demeurent limitées à une perception locale et immédiate. Elles ne fournissent pas d'informations sémantiques détaillées sur la nature des obstacles ni de représentation globale de la scène. Les solutions reposant sur des infrastructures spécifiques ou sur des dispositifs non portables présentent, quant à elles, des contraintes d'accessibilité et de déploiement [5]. Parallèlement, les progrès rapides en vision par ordinateur et en apprentissage profond ont démontré la capacité des systèmes artificiels à analyser des scènes complexes, détecter des objets en temps réel et extraire des informations contextuelles pertinentes.

L'intégration de ces avancées dans des dispositifs embarqués destinés à l'assistance pour la mobilité soulève néanmoins plusieurs défis. Il s'agit de concevoir des architectures capables de percevoir l'environnement de manière fiable malgré les variations d'éclairage, le bruit des capteurs ou les occultations partielles, tout en assurant une estimation de position stable et continue. De plus, les contraintes de calcul et de consommation énergétique imposées par les plateformes portables nécessitent une optimisation rigoureuse des modèles employés. La continuité d'assistance entre environnements outdoor et indoor renforce encore la complexité de cette démarche [6].

Dans cette perspective, l'élaboration d'un système intelligent d'aide à la navigation requiert une approche globale associant perception visuelle, fusion de plusieurs capteurs et stratégies décisionnelles adaptées à l'incertitude. Cette méthode a pour but d'améliorer l'information accessible pour l'utilisateur aveugle, non pas en remplaçant complètement la perception humaine, mais en offrant des repères contextuels appropriés susceptibles de faciliter la navigation et le processus décisionnel.

L'objectif principal de ce travail est de trouver un équilibre entre la performance des algorithmes, la faisabilité intégrée et l'utilité fonctionnelle. Le défi réside dans la conception d'un système cohérent, capable de fonctionner en temps réel, financièrement accessible et flexible pour répondre aux besoins diversifiés des contextes réels. L'idée principale est donc de fournir une architecture unifiée de

Introduction Générale

navigation assistée, qui englobe aussi bien les environnements extérieurs qu'intérieurs, et qui intègre la détection d'obstacles, la classification, l'évitement et la localisation et l'interaction appropriée.

La structure de la thèse reflète cette progression conceptuelle et méthodologique :

- **Le chapitre 1** présente le contexte général de la déficience visuelle et examine les systèmes existants en matière de navigation assistée, en mettant en évidence leurs limites et les défis encore ouverts.
- **Le chapitre 2** présente les fondements théoriques mobilisés, notamment en vision artificielle, apprentissage profond et fusion probabiliste de capteurs.
- **Le chapitre 3** décrit la conception du système de navigation outdoor, en détaillant les modules de détection, de classification et de localisation.
- **Le chapitre 4** est consacré au développement du système indoor et à son implémentation sur plateforme embarquée, en tenant compte des contraintes matérielles et temporelles.
- **Le chapitre 5** présente l'évaluation expérimentale des systèmes proposés, discute les performances obtenues et identifie les perspectives d'amélioration.

À travers cette organisation, la thèse vise à apporter une contribution structurée à la conception de systèmes intelligents d'assistance à la mobilité, en s'appuyant sur une compréhension approfondie du rôle fondamental de la vision dans l'orientation humaine et des défis posés par sa perte.

Chapitre 1 : État de l'Art des Systèmes d'Assistance à la Mobilité pour Personnes Déficiences Visuelles

1.1. Introduction

L'interaction avec l'environnement spatial en situation de déficience visuelle mobilise des mécanismes perceptifs et cognitifs spécifiques, fondés sur l'exploitation d'informations non visuelles et sur des stratégies particulières de représentation de l'espace. Ces mécanismes influencent la manière dont les informations spatiales sont acquises, interprétées et utilisées dans des contextes réels, souvent complexes et dynamiques.

Différentes formes d'assistance ont été développées afin de faciliter l'accès à l'information environnementale et en faciliter l'exploitation. Ces approches se distinguent par les types de données considérées, les modalités de restitution employées et le niveau de traitement appliqué à l'information spatiale. Leur évolution s'inscrit dans un contexte de progrès technologiques continus, permettant une prise en compte de plus en plus fine des caractéristiques de l'environnement.

L'analyse de ces dispositifs soulève des enjeux transversaux liés à la fiabilité des informations produites, à leur cohérence temporelle, à la gestion des situations dynamiques et aux modalités d'évaluation des performances. Ces aspects constituent des éléments clés pour comprendre les limites actuelles et les orientations générales du domaine.

Ce chapitre présente une analyse synthétique des travaux existants relatifs à l'assistance pour la mobilité des personnes déficientes visuelles, visant à mettre en évidence les orientations générales du domaine ainsi que les limites identifiées à l'état actuel des connaissances.

1.2. La Mobilité des Personnes Déficientes Visuelles

1.2.1. Cognition spatiale sans vision

La cognition spatiale en l'absence de vision constitue un champ central pour la compréhension des mécanismes d'orientation et de déplacement des personnes déficientes visuelles. Elle met en évidence la capacité du système nerveux à élaborer des représentations spatiales fonctionnelles sans recours à l'information visuelle, en s'appuyant sur l'intégration coordonnée de modalités sensorielles alternatives, comme illustrée par la figure 1.1. Cette organisation ne relève pas d'un simple phénomène compensatoire, mais résulte d'une adaptation neurofonctionnelle permettant le maintien de capacités spatiales opérationnelles dans des conditions sensorielles contraintes [7].

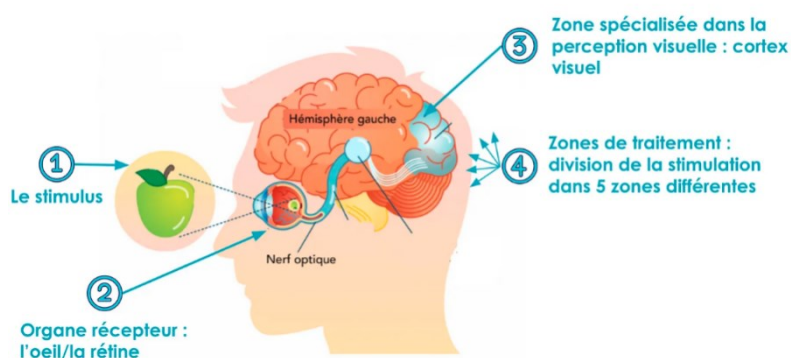


Figure 1.1. La perception spatiale chez l'être humain

1.2.1.1. Intégration sensorielle non visuelle et traitement spatial

En l'absence de vision, le traitement de l'information spatiale repose sur l'exploitation coordonnée des signaux auditifs, tactiles et proprioceptifs. Ces modalités fournissent des indices complémentaires permettant d'estimer la localisation des éléments de l'environnement, leur distance relative et leur organisation spatiale. Les études en neurosciences ont montré que cette intégration repose sur une réorganisation fonctionnelle des aires corticales, dans laquelle certaines régions initialement impliquées dans le traitement visuel participent à l'analyse spatiale d'informations non visuelles. Ce phénomène, souvent qualifié de plasticité cross-modale, constitue un mécanisme central de maintien des capacités spatiales chez les personnes déficientes visuelles [8–10].

L'audition joue un rôle déterminant dans ce cadre. Les différences interaurales de temps et d'intensité, les indices spectraux liés à la morphologie du pavillon de l'oreille ainsi que les informations issues des réflexions sonores permettent d'estimer la direction et la distance des sources. Ces indices sont utilisés non seulement pour localiser des objets ou des personnes, mais également pour inférer la géométrie de l'espace environnant, notamment la présence de surfaces, d'ouvertures ou de couloirs. L'utilisation volontaire de sons auto-générés, tels que les clics de langue ou les bruits produits lors de l'interaction avec le sol, renforce cette analyse en fournissant des informations supplémentaires sur la structure de l'environnement. Des travaux expérimentaux ont mis en évidence que, chez certains utilisateurs entraînés, cette analyse écho-acoustique permet une discrimination spatiale fine, exploitable pour l'orientation et l'évitement d'obstacles [11–13].

Le toucher intervient comme une source d'information spatiale directe, fondée sur l'exploration active de l'environnement. Les informations issues du contact avec les surfaces, les objets et le sol permettent d'identifier des discontinuités, des variations de texture ou des changements de niveau. L'utilisation de la canne blanche étend cette capacité d'exploration en augmentant la portée des interactions tactiles et en fournissant des retours mécaniques et vibratoires exploitables pour l'analyse de l'espace proche. Ces

informations sont intégrées de manière séquentielle et contribuent à l'actualisation continue de la représentation spatiale interne [14].

Les informations proprioceptives et vestibulaires complètent ce processus en assurant une estimation continue de la position et du déplacement du corps dans l'espace. La succession des pas, des rotations et des changements de direction permet de reconstruire la trajectoire parcourue et d'estimer les relations spatiales entre différents lieux. Bien que cette estimation soit sujette à une dérive progressive, elle constitue un support fondamental de l'orientation spatiale, en particulier dans les environnements familiers [15].

1.2.1.2. Mécanismes de mémorisation spatiale sans vision

La mobilité autonome sans vision repose sur la capacité à mémoriser l'environnement de manière suffisamment stable pour permettre la reconnaissance des lieux, la reproduction des trajets et l'anticipation des configurations spatiales. Contrairement à la mémorisation visuelle, qui permet une acquisition globale et immédiate de l'espace, la mémorisation spatiale sans vision s'effectue de manière progressive, séquentielle et distribuée dans le temps. Elle résulte de l'intégration répétée d'informations issues de l'exploration active et de leur consolidation en structures cognitives durables.

Un premier mécanisme fondamental est la mémoire procédurale du déplacement. Les trajets sont initialement mémorisés comme des séquences d'actions successives, associant des mouvements, des changements de direction et des événements perceptifs récurrents. Chaque segment du parcours correspond à une unité fonctionnelle, liée à une action précise, telle qu'un changement d'orientation, une modification du rythme de marche ou une interaction avec un élément de l'environnement. Cette mémoire procédurale permet de reproduire un itinéraire familier sans nécessiter une représentation explicite de l'espace global. Elle constitue souvent la première forme de mémorisation spatiale acquise lors de l'apprentissage d'un nouvel environnement.

À mesure que l'expérience s'accumule, cette mémorisation procédurale est complétée par une mémoire spatiale relationnelle, dans laquelle les différents segments du trajet ne sont plus uniquement chaînés temporellement, mais mis en relation les uns avec les autres. Les lieux sont alors définis par leur position relative au sein du parcours, par leur proximité fonctionnelle ou par leur rôle dans l'organisation du déplacement. Cette transformation permet de réduire la dépendance à un ordre strict d'actions et d'introduire une flexibilité dans la navigation, notamment lors de détours ou de variations de trajet.

La mémorisation spatiale sans vision repose également sur l'identification et la stabilisation de repères spatiaux distinctifs. Ces repères correspondent à des éléments de l'environnement qui présentent une signature perceptive fiable et répétable, permettant leur reconnaissance lors de passages successifs. Leur rôle n'est pas uniquement informatif, mais structurant : ils servent de points d'ancrage pour organiser l'espace environnant et segmenter le trajet en zones cohérentes. La fiabilité temporelle de ces repères

est un facteur déterminant de leur efficacité mémorielle, leur disparition ou leur variabilité pouvant altérer la continuité de la représentation spatiale.

Un autre mécanisme clé est la mémoire de configuration, qui permet d'intégrer plusieurs repères et segments de déplacement en une structure spatiale plus compacte. Cette mémoire ne repose pas sur une reconstruction géométrique précise, mais sur la reconnaissance de configurations spatiales familières, telles que la disposition relative de plusieurs éléments ou la structure typique d'un lieu. Elle facilite l'identification rapide des environnements connus et réduit la charge cognitive nécessaire à l'orientation.

La consolidation de ces différentes formes de mémoire spatiale s'appuie sur la répétition active des déplacements. Chaque parcours contribue à renforcer la stabilité des représentations internes, à corriger les imprécisions accumulées et à améliorer la cohérence globale de l'espace mémorisé. Ce processus est particulièrement sensible à la régularité des trajectoires et à la constance des informations perçues. Les environnements explorés de manière répétée dans des conditions similaires sont mémorisés de façon plus robuste et plus durable.

Enfin, la mémorisation spatiale sans vision intègre une dimension temporelle et contextuelle, permettant d'associer des lieux à des situations, des usages ou des états particuliers de l'environnement. Cette contextualisation facilite la récupération des informations pertinentes en fonction de l'objectif du déplacement et contribue à la stabilité de la navigation à long terme. Elle explique notamment pourquoi certains lieux sont reconnus plus facilement dans des contextes familiers que lors de changements importants de l'environnement sonore ou structurel.

Ainsi, la mémorisation spatiale sans vision ne se réduit ni à une simple accumulation d'informations sensorielles, ni à une reproduction de stratégies visuelles. Elle repose sur un ensemble de mécanismes complémentaires, procéduraux, relationnels et configurationnels qui permettent de construire des représentations spatiales fonctionnelles, adaptatives et exploitables pour la mobilité quotidienne. Ces mécanismes constituent un élément central de la cognition spatiale sans vision et fournissent un cadre conceptuel indispensable pour comprendre les exigences cognitives associées à la navigation autonome des personnes non-voyantes.

1.2.1.3. Principes cognitifs directeurs pour la conception des systèmes d'assistance à la mobilité

Les mécanismes de construction des cartes mentales sans vision mettent en évidence une caractéristique essentielle de la mobilité assistée : l'information spatiale ne peut être exploitée efficacement que si elle s'insère dans une représentation hiérarchisée, séquentielle et orientée vers l'action. Un système d'assistance à la mobilité ne se limite donc pas à fournir des données sur l'environnement, mais intervient directement dans la manière dont ces données sont intégrées aux schémas cognitifs déjà mobilisés par le non-voyant. La pertinence du système dépend alors de sa capacité à enrichir, sans perturber, les structures mentales existantes [16]. Dans cette perspective, l'information transmise doit

dépasser une description métrique brute pour adopter une forme sémantique contextualisée. Identifier la nature d'un objet, sa hauteur ou son comportement dynamique permet d'inscrire immédiatement cet élément dans une stratégie de déplacement cohérente, en réduisant l'effort d'interprétation et en soutenant la prise de décision en temps réel. La hiérarchisation implicite de l'information, distinguant ce qui requiert une réaction immédiate de ce qui relève du repérage spatial, favorise la continuité de la représentation mentale du parcours et limite la fragmentation cognitive [17].

La restitution de l'information s'inscrit dans le prolongement direct des canaux perceptifs mobilisés lors de l'exploration sans vision. Les signaux auditifs spatialisés, lorsqu'ils respectent les conventions naturelles de latéralisation et de distance, s'intègrent aux repères auditifs environnementaux déjà utilisés pour la navigation. De manière complémentaire, les retours haptiques permettent une transmission discrète et continue de l'information spatiale, tout en préservant l'attention auditive pour l'analyse du contexte réel. L'ajustement dynamique des modalités de restitution en fonction de l'environnement et de l'état cognitif de l'utilisateur contribue à maintenir une interaction fluide et soutenable sur la durée. Enfin, l'inscription du système dans les stratégies cognitives de l'utilisateur repose sur une exigence forte de cohérence temporelle et spatiale. Une information tardive, imprécise ou instable perturbe la construction de la carte mentale et fragilise la confiance accordée au dispositif. La performance d'un système d'assistance ne se mesure ainsi pas uniquement à sa capacité de détection, mais à son aptitude pour soutenir la continuité perceptive, l'anticipation des actions et la stabilité des représentations spatiales, conditions indispensables à une mobilité autonome et sécurisée [18].

1.2.1.4. Exploitation cognitive des représentations spatiales chez les personnes non-voyantes

La cognition spatiale chez les personnes non-voyantes ne se limite pas à la simple acquisition d'informations sensorielles ; elle implique un traitement actif et hiérarchisé de ces informations afin de construire et d'exploiter une représentation opérationnelle de l'espace. Cette exploitation permet de maintenir l'orientation, d'anticiper les obstacles et de planifier les déplacements dans des environnements connus ou nouveaux.

Un aspect central de cette exploitation réside dans la mise à jour continue de la position relative. Chaque mouvement effectué par l'individu, déplacement linéaire, rotation ou changement de posture, est intégré avec précision dans la représentation interne de l'espace. Cette mise à jour dynamique permet de corriger les erreurs potentielles issues de l'odométrie corporelle ou des imprécisions proprioceptives, garantissant que la trajectoire mentale reste cohérente avec l'environnement réel.

La correspondance entre perception et mémoire spatiale constitue un autre processus fondamental. Les informations sensorielles reçues en temps réel, qu'elles soient auditives, tactiles ou proprioceptives, sont constamment comparées aux éléments mémorisés de l'environnement. Cette opération cognitive permet de confirmer la localisation de repères connus, de détecter des divergences avec les attentes, et d'ajuster

immédiatement la navigation. Par exemple, un changement inattendu dans l'intensité d'un écho sonore ou une variation tactile du sol peut signaler un obstacle imprévu, ce qui déclenche une adaptation de la trajectoire.

L'organisation hiérarchique et sélective des informations joue un rôle déterminant dans la fluidité du déplacement. Les indices critiques pour la sécurité et la continuité du mouvement sont traités en priorité, tandis que les informations concernant la structure globale de l'environnement sont mobilisées de manière plus stratégique et intégrative. Cette hiérarchisation cognitive permet de limiter la surcharge attentionnelle et de maintenir une navigation efficace, même dans des contextes complexes ou changeants.

L'anticipation spatiale constitue également une caractéristique clé de la cognition non visuelle. Les personnes non-voyantes exploitent leur mémoire spatiale pour anticiper les changements de configuration, la présence d'obstacles ou les transitions entre différentes zones. Cette anticipation repose sur l'apprentissage progressif de l'environnement et sur l'intégration des repères auditifs et tactiles au sein de schémas relationnels stables. Elle permet de préparer mentalement les actions à venir, réduisant ainsi les corrections tardives et favorisant des déplacements fluides et sécurisés.

Enfin, la flexibilité cognitive permet de moduler les stratégies selon le contexte et le niveau de familiarité avec l'environnement. Dans des lieux connus, la navigation s'appuie principalement sur la mémoire spatiale consolidée, tandis que dans des environnements nouveaux ou imprévisibles, l'attention est davantage dirigée vers l'exploitation immédiate des informations sensorielles. Cette capacité d'adaptation garantit une continuité spatiale et temporelle de la navigation, essentielle pour la confiance et l'autonomie de la personne.

Ainsi, l'exploitation cognitive des représentations spatiales chez les personnes non-voyantes repose sur une interaction dynamique entre mémoire, perception et action. Elle permet de transformer les informations sensorielles en connaissance fonctionnelle de l'espace, assurant la sécurité, l'efficacité et la continuité des déplacements, même en l'absence de vision.

1.2.2. Gestion de la navigation selon les degrés de déficience visuelle

La mobilité des personnes déficientes visuelles ne constitue pas une réalité homogène, mais s'inscrit dans un continuum fonctionnel étroitement lié au degré de vision résiduelle. Cette gradation, illustrée par la Figure 1.2, distingue des profils allant de la malvoyance sévère à l'absence totale de perception visuelle. Elle conditionne la manière dont l'information environnementale est perçue, interprétée et intégrée, ainsi que les modes de compensation mis en œuvre au quotidien. Comprendre ces différences fonctionnelles est essentiel pour appréhender la diversité des besoins et des capacités en situation de déplacement [19].

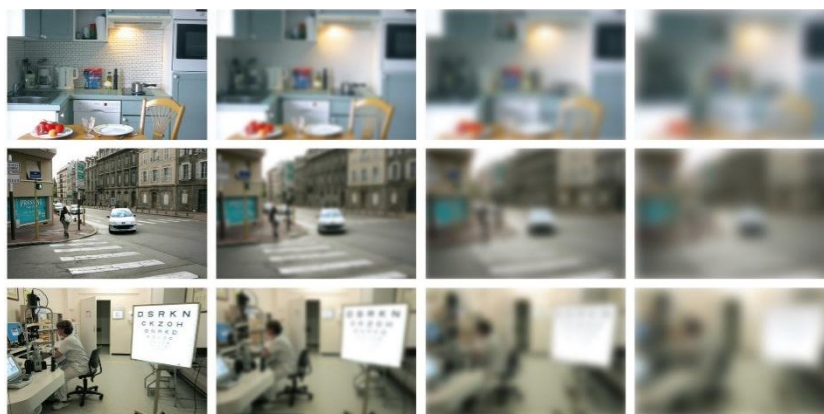


Figure 1.2. Les formes de malvoyances

Chez les personnes disposant d'une vision résiduelle, même fortement altérée, certaines informations visuelles demeurent exploitables. La perception de contrastes grossiers, de variations lumineuses ou de formes globales contribue à structurer l'expérience spatiale. Cette vision partielle n'offre pas une perception continue et détaillée de l'environnement, mais fournit des indices fragmentaires qui participent à l'orientation générale et à l'organisation mentale de l'espace. Ces informations sont systématiquement complétées par des apports auditifs, tactiles et proprioceptifs, donnant lieu à un traitement multisensoriel adaptatif. L'individu développe ainsi une capacité de sélection et de hiérarchisation des informations perçues, privilégiant celles qui présentent la plus forte valeur fonctionnelle.

Lorsque la vision résiduelle devient extrêmement limitée, la perception visuelle perd progressivement son rôle structurant. Les individus concernés peuvent encore percevoir des sensations lumineuses diffuses ou des variations d'intensité, sans que celles-ci ne permettent une identification fiable des éléments de l'environnement. Dans ce contexte, la mobilité repose principalement sur l'exploitation des informations non visuelles et sur la consolidation de schémas perceptifs stables. L'organisation de l'espace s'appuie alors sur des repères auditifs, tactiles et kinesthésiques, associés à une mémorisation précise des configurations spatiales rencontrées. Cette transition vers un fonctionnement majoritairement non visuel implique une mobilisation accrue des capacités attentionnelles et mnésiques, ainsi qu'un apprentissage progressif des environnements fréquentés.

Pour les personnes aveugles innées, l'absence totale d'expérience visuelle conduit à un mode de construction spatiale fondamentalement différent. Les représentations mentales ne reposent pas sur des images visuelles internalisées, mais sur des relations abstraites entre des indices sensoriels non visuels. L'espace est appréhendé comme un ensemble de relations fonctionnelles et séquentielles, organisées à partir de repères auditifs, tactiles et proprioceptifs. Cette structuration relationnelle permet une compréhension fine des configurations spatiales, fondée sur la répétition, l'exploration active et

l'intégration progressive des informations sensorielles disponibles. La cohérence de ces représentations dépend fortement de l'expérience accumulée et de la stabilité des repères utilisés.

Un facteur déterminant dans cette gestion différenciée de la mobilité réside dans l'histoire visuelle de l'individu. Les personnes ayant perdu la vision après une période de fonctionnement visuel conservent souvent des traces mnésiques visuelles qui influencent leur manière d'interpréter l'espace. Ces traces peuvent faciliter l'organisation conceptuelle de l'environnement, en fournissant des cadres de référence issus de l'expérience antérieure. À l'inverse, chez les personnes aveugles congénitales, la construction spatiale est entièrement indépendante de toute référence visuelle, ce qui favorise le développement de stratégies non visuelles optimisées dès les premiers apprentissages.

Cette diversité de profils met en évidence plusieurs principes transversaux dans la gestion de la mobilité selon les degrés de déficience visuelle. Tout d'abord, les modalités sensorielles mobilisées varient en fonction des capacités perceptives disponibles, avec une réallocation progressive des ressources lorsque la vision se dégrade. Ensuite, la structuration de l'espace repose sur des mécanismes de segmentation et de hiérarchisation des informations, indépendamment du canal sensoriel dominant. Enfin, la mémoire spatiale joue un rôle central, sa nature et sa richesse dépendant étroitement de l'expérience sensorielle antérieure et des stratégies d'apprentissage développées [20].

Ainsi, la mobilité des personnes déficientes visuelles ne peut être comprise comme une simple adaptation à une perte sensorielle isolée. Elle résulte d'un équilibre complexe entre perception, cognition et expérience, dans lequel chaque individu développe des modes de fonctionnement spécifiques. Cette variabilité fonctionnelle souligne la nécessité de considérer la diversité des profils visuels dans l'analyse des situations de mobilité quotidienne.

1.2.3. Défis et difficultés de la vie quotidienne des personnes non-voyantes

Au-delà des différences liées aux degrés de déficience visuelle et aux stratégies de navigation qui en découlent, la vie quotidienne des personnes non-voyantes demeure marquée par un ensemble de contraintes structurelles qui influencent directement leur autonomie et leur capacité à se déplacer de manière sûre et efficace. Ces contraintes ne relèvent pas uniquement des capacités perceptives individuelles, mais résultent également de l'organisation des environnements, des modalités d'accès à l'information et des interactions sociales qui accompagnent les situations de mobilité. Le déplacement quotidien constitue ainsi un cadre privilégié pour analyser les difficultés concrètes rencontrées par les personnes non-voyantes, car il mobilise simultanément la perception, la cognition, la prise de décision et l'adaptation à des contextes variés.

L'interaction avec l'environnement physique représente l'un des premiers défis majeurs. Les espaces urbains et architecturaux sont majoritairement conçus sur la base de repères visuels implicites, tels que la signalisation, la disposition spatiale ou les contrastes visuels, qui ne sont pas directement accessibles

aux personnes non-voyantes. Cette dépendance à l'information visuelle rend la perception de l'espace fragmentaire et séquentielle, obligeant l'utilisateur à construire une compréhension progressive de son environnement à partir d'indices non visuels souvent incomplets. Les obstacles fixes, les variations de niveau, les éléments en surplomb ou les changements de configuration spatiale constituent des sources de risque récurrentes, en particulier dans des environnements complexes ou peu structurés.

La dynamique environnementale accentue ces difficultés. Les espaces de mobilité quotidienne, tels que les rues, les transports ou les lieux publics, sont caractérisés par une variabilité constante liée à la présence d'autres usagers, de véhicules, d'objets temporaires ou de modifications imprévues de l'environnement. Cette variabilité impose une capacité d'adaptation rapide et continue, tout en sollicitant fortement les ressources attentionnelles. L'absence de vision directe rend la détection précoce des changements plus difficile, ce qui peut entraîner des situations d'incertitude ou de désorientation, notamment dans les contextes à forte densité ou à forte activité.

L'accès à l'information spatiale et fonctionnelle constitue un autre enjeu central. De nombreux éléments essentiels à la navigation, tels que les indications directionnelles, les numéros de rues, les plans de bâtiments ou les informations de transport, sont principalement transmis sous forme visuelle. Cette modalité d'information limite l'autonomie des personnes non-voyantes, qui doivent souvent recourir à des stratégies de compensation ou à l'assistance d'autrui pour interpréter leur environnement. Cette dépendance informationnelle peut ralentir les déplacements, augmenter la charge cognitive et réduire la capacité à anticiper les actions à entreprendre.

Les interactions sociales en situation de mobilité représentent également une source de difficulté spécifique. Les personnes non-voyantes sont fréquemment confrontées à des comportements inadaptés, tels que des aides non sollicitées, des attitudes paternalistes ou, à l'inverse, une absence totale d'assistance lorsque celle-ci serait nécessaire. Ces interactions peuvent perturber la continuité du déplacement, générer un sentiment d'insécurité ou limiter l'initiative individuelle. De plus, les normes sociales implicites liées à la circulation, aux priorités de passage ou à l'usage des espaces partagés reposent souvent sur des signaux visuels, ce qui complique leur interprétation dans un contexte non visuel.

Les environnements non familiers constituent un défi particulièrement marqué. Contrairement aux espaces connus, dans lesquels des repères mémorisés peuvent être mobilisés, les environnements nouveaux exigent une exploration progressive et une construction active de la représentation spatiale. Cette phase d'exploration est souvent coûteuse en termes de temps et d'effort cognitif, et peut être associée à un stress accru lié à l'incertitude et au risque perçu. La difficulté à anticiper la structure globale de l'espace limite la capacité à planifier les déplacements et à choisir des itinéraires optimaux.

À ces contraintes s'ajoutent des facteurs liés aux conditions sensorielles et contextuelles. Le bruit ambiant, les conditions météorologiques ou la qualité acoustique des espaces peuvent altérer la perception des indices auditifs utilisés pour l'orientation et la détection d'obstacles. De même, les environnements contraints ou dégradés, tels que les zones en travaux, les espaces encombrés ou les situations d'urgence, accentuent les limites des stratégies de navigation non visuelles et augmentent le risque d'erreurs ou de situations dangereuses.

Les conséquences de ces difficultés dépassent le cadre strict de la mobilité. Elles influencent l'accès à l'éducation, à l'emploi et aux activités sociales, en conditionnant la capacité à se déplacer de manière autonome et fiable. Les contraintes liées aux déplacements quotidiens peuvent restreindre les choix d'activités, limiter la spontanéité et renforcer la dépendance à l'égard de tiers ou de dispositifs d'assistance [21]. Cette réalité souligne que les défis de la mobilité ne sont pas uniquement techniques, mais s'inscrivent dans une problématique plus large d'inclusion et de participation sociale.

L'ensemble de ces éléments met en évidence que les difficultés rencontrées par les personnes non-voyantes dans leur vie quotidienne ne peuvent être réduites à une simple absence de vision. Elles résultent de l'interaction entre des environnements majoritairement conçus pour des usagers voyants, des contraintes informationnelles persistantes et des exigences élevées en matière d'adaptation et de prise de décision. Face à ces défis, les personnes non-voyantes ont historiquement développé et adopté des aides spécifiques destinées à sécuriser et faciliter leurs déplacements. L'analyse de ces aides, de leurs apports et de leurs limites, constitue une étape essentielle pour comprendre l'évolution des solutions d'assistance à la mobilité, et justifie l'examen des dispositifs traditionnels avant l'émergence des approches technologiques plus avancées.

1.3. Aides traditionnelles et leurs limites

L'analyse des stratégies de navigation des personnes déficientes visuelles révèle des compétences d'adaptation remarquables, mais également des contraintes liées à la portée et à la nature des informations sensorielles disponibles. Les aides traditionnelles, principalement la canne blanche et le chien guide, constituent des solutions éprouvées offrant un support essentiel à la mobilité. Cependant, une étude détaillée de leurs mécanismes et de leurs limites met en évidence des besoins non satisfaits, particulièrement dans les environnements urbains contemporains complexes.

1.3.1. La canne blanche

La canne blanche, perfectionnée au XX^e siècle, constitue un dispositif haptique fondamental permettant à l'utilisateur de transformer un geste simple, le balayage rythmique, en un flux continu d'informations spatiales. Chaque contact avec le sol transmet simultanément trois types d'informations : la position, la texture et la résistance. Cette interaction directe enrichit la représentation mentale de l'environnement

et constitue un outil d'anticipation des obstacles immédiats [22]. La figure 1.3 présente la canne blanche inventée par James Biggs en 1921.

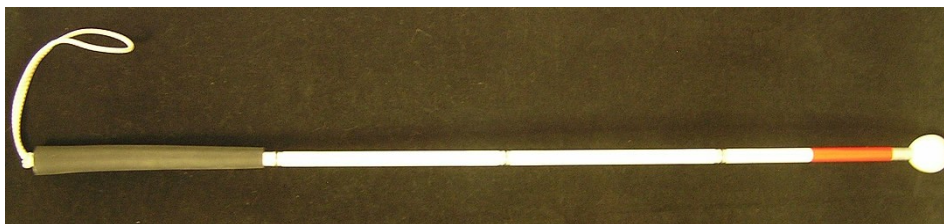


Figure 1.3. Premier outil pour la navigation pour les non-voyants

Cependant, la canne présente des limites inhérentes à sa conception :

- Exploration séquentielle : La navigation avec la canne est linéaire, traitant l'espace segment par segment. Cette exploration fragmentaire impose une reconstruction cognitive d'une image globale à partir d'indices partiels, augmentant la charge de traitement et l'effort attentif.
- Portée limitée : La canne détecte les obstacles sur une distance maximale d'environ 1,5 mètres, créant une « zone de certitude » restreinte. Cette limitation est critique face aux obstacles mobiles rapides (piétons, cyclistes, véhicules), où le temps de réaction peut être insuffisant pour éviter un incident.
- Absence de contextualisation : La canne informe sur la présence d'un obstacle, mais pas sur sa nature ou son comportement. Cette « cécité sémantique » oblige l'utilisateur à interpréter l'environnement en temps réel et à élaborer des hypothèses sur les obstacles rencontrés, augmentant l'incertitude et la charge décisionnelle.

1.3.2. Le chien guide : Une intelligence biologique

Le chien guide représente une approche complémentaire, introduisant une forme d'intelligence autonome capable d'interpréter et d'anticiper les obstacles. L'entraînement spécialisé des chiens guides permet de créer une unité cognitive binôme homme-chien, où certaines décisions de navigation sont partagées, améliorant la sécurité et la fluidité du déplacement [23]. Malgré ces avantages, des limites techniques et fonctionnelles persistent :

- Granularité de l'information : Les signaux transmis par le chien sont abstraits et limités à des indications binaires ou qualitatives ("obstacle présent", "danger imminent"). Ils ne fournissent pas de détails sur la nature ou la position exacte de l'obstacle, essentiels pour planifier une trajectoire optimale.

- Variabilité biologique : L'attention et la performance du chien dépendent de son état physiologique et émotionnel, ce qui introduit une incertitude supplémentaire pour l'utilisateur.
- Inadaptation aux environnements numériques : Les chiens guides ne peuvent interagir avec des dispositifs numériques, des signalétiques électroniques ou des systèmes de transport intelligents, de plus en plus présents dans les espaces urbains modernes.

1.3.3. Insuffisances fonctionnelles des aides traditionnelles et émergence de nouveaux besoins d'assistance

L'analyse approfondie de la canne blanche et du chien guide met en évidence leur rôle central dans la mobilité des personnes déficientes visuelles, tout en révélant des insuffisances fonctionnelles structurelles liées à leur nature même. Ces dispositifs reposent sur des principes éprouvés, interaction haptique directe pour la canne, interprétation biologique pour le chien guide, qui ont démontré leur efficacité dans des contextes variés. Toutefois, l'évolution des environnements de mobilité, marquée par une densification spatiale, une accélération des flux et une hybridation croissante entre espaces physiques et numériques, accentue les écarts entre les capacités offertes par ces aides et les exigences actuelles de déplacement autonome.

Un premier besoin majeur concerne l'extension de l'horizon informationnel disponible pour l'utilisateur. La canne blanche, par sa portée limitée, et le chien guide, par la nature implicite de ses signaux, fournissent une information principalement locale et immédiate. Cette restriction réduit la capacité d'anticipation et contraint l'utilisateur à opérer dans un cadre temporel étroit, où les décisions doivent être prises sous pression. Dans des environnements urbains caractérisés par des interactions rapides et des changements fréquents, cette absence de visibilité prospective constitue une limite déterminante. Il apparaît ainsi nécessaire de disposer de dispositifs capables de fournir une information étendue dans l'espace et suffisamment en amont pour soutenir une planification plus stable et moins réactive du déplacement.

Au-delà de la portée, les aides traditionnelles se distinguent par une faible explicitation du contenu informationnel. La canne signale une présence physique sans en qualifier la nature, tandis que le chien guide transmet des indications globales, issues de son propre jugement, sans accès direct aux critères sous-jacents à la décision. Cette absence de description explicite impose à l'utilisateur une interprétation permanente, fondée sur l'expérience et le raisonnement inférentiel. Il en résulte une charge cognitive accrue, en particulier dans les situations complexes ou inédites. L'émergence d'un besoin de qualification fonctionnelle de l'environnement se dessine alors clairement : l'assistance ne peut se

limiter à signaler, elle doit permettre de comprendre, au moins partiellement, les éléments rencontrés et leur rôle dans la situation de déplacement.

Par ailleurs, les aides traditionnelles opèrent dans un cadre essentiellement déconnecté des systèmes d'information contemporains. La canne blanche et le chien guide interagissent exclusivement avec l'environnement physique immédiat, sans accès aux données issues des infrastructures numériques, des dispositifs connectés ou des services urbains intelligents. Or, ces sources d'information constituent désormais une composante structurante de l'espace public, influençant l'orientation, l'accès aux services et la gestion des flux. L'incapacité des aides classiques à exploiter ces ressources limite leur potentiel d'assistance et accentue la dépendance de l'utilisateur à des aides humaines externes ou à une connaissance préalable des lieux.

Un autre enjeu réside dans la gestion de la variabilité et de l'évolution des situations de mobilité. Les environnements réels sont soumis à des modifications fréquentes, qu'elles soient temporaires ou durables. Les aides traditionnelles, fondées sur une interaction directe et instantanée, n'intègrent aucun mécanisme de mise à jour ou d'adaptation structurelle à ces changements. Cette rigidité fonctionnelle expose l'utilisateur à des situations imprévues, nécessitant un effort constant d'adaptation individuelle. L'émergence d'un besoin d'assistance capable d'intégrer la dynamique des environnements et de maintenir une cohérence informationnelle dans le temps devient alors manifeste.

Enfin, l'ensemble de ces limites souligne un besoin transversal de maîtrise de la complexité informationnelle. Toute extension des capacités d'assistance doit éviter de transférer la complexité du système vers l'utilisateur. Les aides traditionnelles, bien que limitées, présentent l'avantage d'une simplicité d'usage et d'une interprétation relativement intuitive. Les besoins émergents ne concernent donc pas uniquement l'enrichissement de l'information, mais également sa structuration, sa hiérarchisation et son adéquation aux capacités de traitement de l'utilisateur, afin de préserver la fluidité et la sécurité du déplacement.

Ainsi, les insuffisances observées dans les aides traditionnelles ne remettent pas en cause leur pertinence, mais révèlent les contours d'un ensemble de besoins fonctionnels nouveaux. Ces besoins, liés à l'anticipation, à la qualification de l'information, à l'intégration des environnements numériques, à l'adaptation contextuelle et à la gestion de la complexité, définissent le cadre conceptuel à partir duquel se développent les évolutions technologiques ultérieures. Ils marquent une transition progressive entre des dispositifs fondés sur une interaction directe avec l'environnement et des systèmes d'assistance plus avancés, capables d'enrichir la perception et le soutien à la mobilité sans se substituer à l'utilisateur.

1.4. Évolution des Systèmes d'Assistance Technologiques : Des Capteurs à l'IA

La reconnaissance des limites cognitives et pratiques des aides traditionnelles a catalysé une évolution technologique en trois vagues successives. Cette progression, cumulative plutôt que strictement linéaire, a abordé tour à tour des défis spécifiques tout en révélant de nouvelles frontières. Elle illustre le passage d'une approche purement métrique « mesurer des distances » à une approche cognitive véritablement intégrée « comprendre l'environnement pour assister la décision ».

1.4.1. Première Génération : L'ère des Sens Électroniques Élémentaires

La première génération, apparue dans les années 1970-1980, visait à transcender les limites du toucher et de l'audition par la création de « sens artificiels », permettant de percevoir l'environnement à distance. L'objectif n'était pas encore de comprendre le monde, mais de le mesurer avec une précision supérieure aux capacités biologiques restantes, jetant les bases techniques et conceptuelles de l'assistance électronique moderne.

1.4.1.1. L'Écholocalisation Ultrasonore

Inspiré de l'écholocalisation biologique, ce paradigme reposait sur l'émission d'impulsions ultrasonores (40–200 kHz) et la mesure de leur temps de propagation aller-retour [24-25]. Le Sonic Pathfinder [26], porté sur un bandeau céphalique, utilisait cinq transducteurs (deux émetteurs et trois récepteurs) et un micro-ordinateur embarqué pour estimer la distance et la position latérale des obstacles, avec retour binaural. Le Sonicguide [27] transformait les distances en sons stéréophoniques, offrant une représentation auditive complète, mais provoquant une surcharge cognitive. Enfin, le Russell Pathsounder [28], plus simple, se limitait à un transducteur et un retour vibratoire, privilégiant fiabilité et intuitivité sur la richesse informationnelle.

Ces systèmes étaient limités par l'atténuation atmosphérique exponentielle : à 5 mètres, un signal à 40 kHz perdait déjà la moitié de son amplitude (≈ 6 dB), réduisant leur portée pratique et limitant l'anticipation des obstacles mobiles.

1.4.1.2. La triangulation Optique Infrarouge (Mesurer par la Lumière)

Parallèlement, les dispositifs IR utilisaient un faisceau modulé et la mesure de l'angle réfléchi pour obtenir une distance précise [29-30]. Le TomPouce [31] comme illustré dans la figure 1.4, employait des diodes GaAs à 880 nm et un phototransistor modulé à 455 kHz, avec un retour vibratoire croissant selon la proximité. Le Mowat Sensor [32] utilisait une photodiode avalanche pour amplifier le signal, mais restait sensible aux interférences lumineuses et saturations solaires. Ces systèmes, bien que précis, étaient limités à 2–3 mètres, sensibles à l'éclairage ambiant et incapables de détecter des surfaces absorbantes.



Figure 1.4. Le TomPouce

1.4.1.3. Laser par Temps de Vol

Enfin, les prototypes laser, comme le LaserCane [33], mesuraient le temps de vol d'impulsions laser pour une détection centimétrique. Trois diodes laser intégrées dans une canne traditionnelle permettaient de détecter le sol, les obstacles frontaux et les obstacles en hauteur. La complexité mécanique, la consommation énergétique et la faible puissance laser limitaient toutefois la portée et la robustesse.

Malgré leurs différences techniques, ces trois paradigmes partageaient une lacune majeure : l'absence de contextualisation sémantique. Ils répondaient à la question « à quelle distance ? », mais pas « qu'est-ce que c'est ? » ou « est-ce dangereux ? ». Cette première génération fut ainsi un laboratoire technologique et cognitif, démontrant la faisabilité de l'extension artificielle de la perception, tout en soulignant l'importance de la fiabilité, de l'ergonomie cognitive et de la complémentarité avec les aides traditionnelles.

1.4.2. Deuxième Génération (1995–2010) : Navigation Géoréférencée et Cartographie Numérique

Les limites des systèmes de première génération ont conduit à un pivot stratégique : la navigation à grande échelle. L'amélioration du GPS civil après 2000 permit de se concentrer sur la localisation absolue et la planification d'itinéraires, plutôt que sur la simple détection locale d'obstacles [34].

1.4.2.1. Le GPS et Trilatération Temporelle

Le GPS détermine la position absolue via la mesure du temps de propagation des signaux satellites, fournissant une référence pour aligner l'utilisateur avec des cartes numériques. Le Personal Guidance System (PGS) [35] exploitait un GPS différentiel (DGPS) [36] et un guidage auditif binaural pour indiquer la direction et la distance de la destination. L'Audio GPS [37] utilisait des clics spatialisés, tandis que des systèmes hybrides [38] faisaient intervenir un opérateur humain pour transmettre les instructions. Ces travaux ont démontré que la qualité cognitive de l'interface était aussi cruciale que la précision GPS.

1.4.2.2. PDA et des Systèmes Intégrés

Les prototypes universitaires ont évolué vers des systèmes intégrés comme Drishti [39], combinant GPS différentiel et navigation inertielle avec des balises RFID pour continuité indoor/outdoor. Audio Browser [40] et Trekker [41] ont amélioré l'ergonomie auditive et l'appropriation par l'utilisateur. L'intégration tactile et audio, comme Timbremap [42], a permis d'explorer la carte par le toucher avec retour sonore, améliorant l'accessibilité et réduisant la charge cognitive.

1.4.2.2. La Transition Vers le Smartphone et les Applications Modernes

Entre 2007 et 2010, les smartphones ont intégré des capteurs et du calcul, permettant des solutions GPS, Bluetooth et vision pour le wayfinding. Les travaux de Mulloni et al. [43] et Coughlan & Manduchi [44] ont démontré l'usage de la caméra pour localiser et identifier des obstacles et marquages au sol. Malgré ces avancées, la dichotomie entre la précision de la carte et l'opacité de l'environnement immédiat demeure un obstacle majeur, révélant la nécessité de systèmes capables de « voir » et de comprendre la scène en temps réel.

1.4.3. Troisième Génération (2010-Présent) : Perception Intelligente et de la Compréhension Sémantique

L'échec fondamental de la deuxième génération à percevoir l'environnement immédiat, malgré sa maîtrise de la navigation stratégique, a créé une impasse cognitive nécessitant une rupture conceptuelle. Cette troisième génération, amorcée autour de 2010, se caractérise par un objectif radicalement nouveau : passer de la simple localisation et du guidage à la compréhension contextuelle en temps réel de l'environnement. Cette ambition est rendue possible grâce à la convergence de plusieurs avancées : la miniaturisation et la généralisation de capteurs riches tels que les caméras RGB-D et les LiDARs compacts, l'explosion de la puissance de calcul embarquée et les progrès de l'apprentissage profond (Deep Learning) en vision par ordinateur et intelligence artificielle. Le smartphone et les dispositifs portables deviennent ainsi de véritables plateformes de perception cognitive mobiles, capables d'analyser le flux visuel et de restituer des informations sémantiques en temps réel, à l'image d'un système visuel biologique mais avec des capacités de traitement et de reconnaissance supérieures.

1.4.3.1. Émergence des Systèmes de Vision Assistée

Les premiers travaux de cette génération ont exploité les capteurs de profondeur émergents, comme le Kinect, et les algorithmes fondamentaux de vision par ordinateur pour créer des systèmes de perception 3D, posant les bases techniques de la détection d'obstacles et de la compréhension géométrique de l'environnement [45]. Parmi ces travaux, Bousbia-Salah et al. [46] ont développé un dispositif intégrant trois capteurs à ultrasons, deux portés à l'épaule et un sur la canne, capables de détecter les obstacles au sol et en hauteur jusqu'à six mètres, avec un retour haptique via des vibrateurs, illustrant l'efficacité

d'une boucle perception-action en temps réel pour sécuriser les déplacements urbains. Aladren et al. [47] ont ensuite proposé un système utilisant un capteur RGB-D avec une « expansion de plage » algorithmique pour extrapoler la carte de profondeur au-delà du champ limité du capteur, anticipant ainsi la détection précoce des obstacles. Le problème de la localisation précise en intérieur a également motivé l'adaptation des algorithmes de SLAM, comme le montrent les travaux de Nguyen et al. [48], qui ont appliqué le SLAM visuel sur un robot mobile pour assister les personnes déficientes visuelles, construisant simultanément une carte de l'environnement et localisant l'utilisateur. Ces recherches ont identifié les défis liés à la variabilité d'éclairage, aux angles de vue et à la reconnaissance des objets du quotidien.

1.4.3.2. Intégration de l'apprentissage profond et optimisation systémique des architectures

L'avènement des réseaux de neurones convolutifs (CNN) au milieu des années 2010 a permis de passer de la détection manuelle et fragile d'objets à des modèles capables de reconnaître des centaines de catégories avec robustesse. Tapu et al. [49] ont illustré cette transition avec le système DEEP-SEE, capable de détecter obstacles, textes et feux de signalisation à partir d'un simple smartphone, extrayant ainsi une information sémantique riche. Parallèlement, le SLAM a gagné en robustesse : Endo et al. [50] ont appliqué le LSD-SLAM pour générer des cartes denses à partir d'images, tandis que Zhang et Ye [51] ont renforcé la cohérence de la localisation intérieure en intégrant la reconnaissance de structures fixes comme murs et portes. L'optimisation pour des dispositifs mobiles a été explorée par Wong et al. [52], qui ont utilisé MobileNetV2-SSD pour concevoir un système de détection embarqué, et par Alhichri et al. [53], qui ont proposé l'étiquetage multi-label avec SqueezeNet pour permettre la reconnaissance simultanée de plusieurs objets dans une scène. Ces approches ont maximisé la précision et la vitesse tout en respectant les contraintes de mémoire et de puissance de calcul, rendant possible une détection en temps réel sur smartphone ou dispositif portable.

1.4.3.3. Intégration Systémique, Fusion Multi-Capteurs et IA Efficace

À partir de 2020, la recherche s'est orientée vers la création de systèmes holistiques et robustes, combinant perception, localisation, décision et interface. Le système V-Eye [54] illustre cette tendance, en intégrant la détection sémantique des objets (personnes, véhicules, obstacles) et l'estimation de trajectoire 3D pour une navigation anticipative et sécurisée. Vaishnav et al. [55] et Afif et al. [56] ont développé des dispositifs portables combinant caméra et unité de calcul, offrant discrétion et liberté de mouvement par rapport au smartphone. L'intégration profonde de la sémantique et de la géométrie permet aux systèmes modernes de distinguer, par exemple, une « chaise vide » d'une « personne assise » et d'évaluer si cette dernière se trouve sur la trajectoire, offrant une navigation sociale et sécurisée. Les architectures les plus récentes [57] orchestrent la perception multi-capteurs [58] (vision, LiDAR, IMU, GPS), la localisation hybride [59] (SLAM visuel/inertiel et fusion GPS), la compréhension de scène via Vision Transformers comme DINOv2 et la génération d'instructions adaptatives [60].

L'objectif est de passer des prototypes de laboratoire à des systèmes portables fiables, capables de fonctionner dans des environnements complexes et dynamiques, tout en détectant de petits objets et en classifiant les scènes de manière dynamique. Cette troisième génération représente ainsi la maturité des systèmes d'assistance, permettant non seulement de localiser et guider l'utilisateur, mais également de percevoir, comprendre et agir dans le monde réel avec une autonomie proche de celle procurée par la vision humaine.

1.5. Critères d'évaluation des systèmes d'assistance à la mobilité

L'évaluation des systèmes d'assistance à la mobilité constitue un enjeu fondamental pour apprécier leur capacité à soutenir un déplacement autonome, sûr et fluide dans des environnements variés. Ces systèmes s'inscrivent dans des situations d'usage complexes, caractérisées par l'interaction simultanée entre des dispositifs techniques, des capacités humaines et des contraintes environnementales dynamiques. Cette complexité impose une approche d'évaluation multidimensionnelle, dépassant une analyse strictement technologique ou instrumentale.

Les systèmes d'assistance à la mobilité ne se limitent pas à produire des résultats techniques mesurables ; ils participent directement à l'activité de déplacement et influencent les stratégies d'action, d'anticipation et de prise de décision. Leur évaluation doit ainsi prendre en compte la manière dont les informations produites sont intégrées dans le comportement de l'utilisateur, ainsi que leur impact sur la sécurité, le confort et la continuité du déplacement.

1.5.1. Performance fonctionnelle globale

La performance fonctionnelle constitue un premier axe d'analyse essentiel. Elle vise à caractériser la capacité du système à assurer les fonctions pour lesquelles il a été conçu, dans des conditions d'utilisation données. Cette performance ne peut être réduite à des mesures ponctuelles d'efficacité ou d'exactitude, mais doit être envisagée dans une perspective globale et continue.

Dans le contexte de la mobilité assistée, la performance fonctionnelle intègre la capacité du système à fournir des informations pertinentes au moment opportun, à maintenir un fonctionnement stable sur la durée et à préserver une continuité de service malgré les variations de contexte. Un système peut présenter des performances élevées dans des conditions favorables tout en devenant peu fiable dès que l'environnement se complexifie, ce qui souligne l'importance d'une analyse étendue des conditions d'usage.

La notion de robustesse fonctionnelle est étroitement liée à cette performance globale. Elle renvoie à la capacité du système à tolérer des perturbations telles que le bruit sensoriel, les changements d'éclairage, les variations spatiales ou les comportements imprévus de l'utilisateur. Une robustesse insuffisante peut

entraîner des interruptions de service ou des informations incohérentes, compromettant la continuité du déplacement et la confiance accordée au dispositif.

1.5.2. Précision, fiabilité et cohérence temporelle

La précision des informations délivrées constitue un critère central dans l'évaluation des systèmes d'assistance à la mobilité. Elle conditionne la capacité de l'utilisateur à interpréter correctement la situation et à prendre des décisions adaptées. Toutefois, la précision ne peut être considérée isolément, car elle n'est pleinement pertinente que si elle s'inscrit dans un cadre fiable et cohérent dans le temps.

La fiabilité renvoie à la constance des performances du système et à sa capacité de produire des informations correctes de manière répétée. Une information précise mais fournie de façon intermittente ou imprévisible peut s'avérer moins utile qu'une information légèrement moins précise mais stable. La cohérence temporelle, entendue comme la stabilité des sorties du système sur des intervalles courts, joue un rôle clé dans la compréhension et l'anticipation.

Les systèmes d'assistance à la mobilité doivent également gérer les situations d'incertitude. La capacité à identifier des conditions ambiguës et à éviter des restitutions trompeuses constitue un élément déterminant de l'évaluation. Une restitution non fiable peut induire une fausse confiance et exposer l'utilisateur à des situations à risque. À l'inverse, la prise en compte explicite des limites du système favorise une interaction plus prudente et plus adaptée aux contraintes réelles.

1.5.3. Réactivité et latence

La réactivité du système correspond à sa capacité de produire des informations exploitables dans des délais compatibles avec la dynamique du déplacement. En situation de mobilité, les décisions doivent souvent être prises dans des fenêtres temporelles réduites, notamment lors de l'évitement d'obstacles ou de changements de trajectoire.

La latence, définie comme le délai entre l'acquisition de l'information et sa restitution perceptible, constitue un facteur critique. Elle résulte de l'ensemble des étapes de la chaîne de traitement, incluant la captation des données, leur traitement, leur interprétation et leur restitution. Une latence excessive peut rendre l'information obsolète, voire contre-productive, en induisant des réactions tardives ou inappropriées.

Au-delà de la valeur moyenne de la latence, sa variabilité représente un enjeu majeur. Des délais fluctuants peuvent perturber la synchronisation entre l'action humaine et la réponse du système, augmentant la charge cognitive et réduisant la fluidité du déplacement. L'évaluation de la réactivité doit donc considérer à la fois l'amplitude et la stabilité des délais de restitution.

1.5.4. Charge cognitive et clarté informationnelle

La charge cognitive induite par l'utilisation d'un système d'assistance constitue un critère central dans l'analyse de son efficacité globale. Le déplacement autonome mobilise déjà des ressources attentionnelles importantes, notamment pour l'orientation, l'anticipation et la gestion des interactions avec l'environnement. Une assistance mal conçue peut accentuer cette charge au lieu de la réduire.

L'évaluation de la charge cognitive repose sur différents indicateurs, tels que le temps de réaction, le taux d'erreur, la vitesse de déplacement ou des mesures subjectives de confort mental. Une surcharge cognitive peut résulter d'un excès d'informations, d'une hiérarchisation insuffisante ou d'une restitution difficile à interpréter dans le flux de l'action.

La clarté informationnelle joue un rôle déterminant dans la réduction de cette charge. Une information claire, structurée et cohérente favorise une intégration rapide et quasi automatique, limitant l'effort d'interprétation. À l'inverse, une restitution ambiguë ou instable peut monopoliser l'attention et détourner l'utilisateur de la tâche principale de déplacement.

1.5.5. Sécurité et maîtrise des situations critiques

La sécurité constitue un critère prioritaire dans l'évaluation des systèmes d'assistance à la mobilité. Elle concerne la capacité du dispositif à réduire les risques d'accident, tels que les collisions, les chutes ou la désorientation, tout en évitant l'introduction de nouveaux dangers liés à son fonctionnement.

L'analyse de la sécurité ne se limite pas à l'absence d'incidents observés dans des conditions contrôlées. Elle doit également prendre en compte la gestion des situations critiques, la dégradation progressive des performances et le comportement du système en cas de défaillance partielle ou totale. La capacité à maintenir un fonctionnement minimal ou à éviter des restitutions trompeuses en situation dégradée constitue un élément central.

La sécurité perçue représente également un aspect déterminant. Un système peut être techniquement sûr tout en générant un sentiment d'insécurité s'il est imprévisible ou difficile à interpréter. Cette perception influence fortement l'acceptation du dispositif et la manière dont il est utilisé dans la durée.

1.5.6. Acceptabilité et confort d'usage

L'acceptabilité conditionne l'adoption effective des systèmes d'assistance à la mobilité. Elle dépend d'un ensemble de facteurs liés au confort physique, à la discrétion du dispositif et à sa compatibilité avec les habitudes de déplacement. Un système performant sur le plan technique peut être peu utilisé s'il est perçu comme contraignant ou intrusif.

Le confort d'usage ne se limite pas à l'ergonomie matérielle, mais inclut également le confort cognitif et perceptif. Une assistance fatigante, sollicitant excessivement l'attention ou générant une tension prolongée, risque d'être réservée à des situations ponctuelles, limitant son impact réel sur l'autonomie.

L'usage prolongé constitue un indicateur important de l'acceptabilité. Un dispositif véritablement adapté doit pouvoir être intégré durablement dans les pratiques quotidiennes sans engendrer de lassitude ou de rejet progressif.

1.5.7. Adaptabilité et capacité de généralisation

La capacité d'un système à s'adapter à une diversité de situations constitue un critère majeur d'évaluation. Les environnements de mobilité sont par nature variables, tant sur le plan spatial que sur le plan contextuel. Un système trop spécialisé peut présenter des performances élevées dans un cadre restreint, mais devenir inopérant dès que les conditions évoluent.

L'adaptabilité concerne la flexibilité des paramètres, la prise en compte de profils utilisateurs variés et la capacité à fonctionner dans des environnements hétérogènes. Elle conditionne la généralisation des performances et le potentiel de déploiement à plus grande échelle.

La généralisation ne se réduit pas à la reproductibilité des résultats, mais implique une stabilité des performances face à des usages non anticipés. Elle constitue un indicateur de maturité et de robustesse du système dans des contextes réels.

1.5.8. Cadres d'évaluation et limites méthodologiques

Les approches d'évaluation des systèmes d'assistance à la mobilité présentent des limites inhérentes à la complexité des situations étudiées. Les évaluations en environnement contrôlé permettent une maîtrise fine des paramètres, mais peinent à reproduire la richesse et l'imprévisibilité des situations réelles.

À l'inverse, les évaluations en conditions naturelles offrent une validité contextuelle élevée, au prix d'une variabilité importante et d'une difficulté accrue à isoler les facteurs explicatifs. Le choix du cadre d'évaluation influence directement les résultats obtenus et leur interprétation, ce qui complique les comparaisons entre systèmes.

Ces limites soulignent la nécessité d'approches combinant différentes échelles d'analyse, afin de mieux appréhender la performance globale et les usages effectifs.

Ces limites soulignent la nécessité d'approches d'évaluation hybrides, combinant des tests contrôlés et des études de terrain, afin de fournir une image plus complète des performances et des usages.

1.5.9. Enjeux actuels de l'évaluation des systèmes d'assistance

Les évolutions récentes des technologies d'assistance renforcent la nécessité de critères d'évaluation intégratifs, capables de refléter la complexité des interactions entre l'utilisateur, le système et l'environnement. L'évaluation tend à s'orienter vers une prise en compte globale de l'expérience de déplacement, plutôt que vers des métriques isolées.

L'émergence de systèmes adaptatifs pose également la question de l'évaluation dynamique, tenant compte de l'évolution conjointe du comportement du dispositif et des stratégies de l'utilisateur. Ces enjeux constituent un axe central pour l'analyse et la compréhension des systèmes d'assistance à la mobilité et conditionnent leur évolution vers des usages plus robustes et plus durables.

1.6. Conclusion

Ce chapitre a proposé une vue d'ensemble des systèmes d'assistance à la mobilité destinés aux personnes non-voyantes, en s'appuyant sur une analyse de la navigation sans vision et des mécanismes par lesquels l'espace est perçu et exploité en l'absence du canal visuel. L'étude de la cognition spatiale non visuelle a permis de dégager une représentation fonctionnelle du monde fondée sur l'intégration sensorielle, la mémorisation spatiale et l'exploitation de relations environnementales.

L'examen des aides traditionnelles a mis en évidence leurs apports essentiels ainsi que leurs limites, soulignant l'émergence de besoins fonctionnels non satisfaits et la nécessité d'approches d'assistance plus intégrées. L'évolution des systèmes technologiques, depuis les capteurs élémentaires jusqu'aux architectures intégrant la perception intelligente et la compréhension sémantique, illustre une progression vers des dispositifs capables de fournir une information plus riche et mieux contextualisée.

La structuration des approches selon des sous-systèmes fonctionnels, détection, évitement, localisation et classification, a permis de mettre en évidence la diversité des solutions proposées et les compromis associés. L'introduction de critères d'évaluation multidimensionnels offre enfin un cadre permettant d'appréhender ces systèmes dans leur globalité, en tenant compte de leurs performances, de leur comportement temporel et de leur capacité d'adaptation.

Dans cette continuité, le chapitre suivant est consacré aux fondements théoriques et à la modélisation générale des systèmes d'assistance à la mobilité, afin de structurer formellement les architectures et les mécanismes décisionnels sous-jacents.

Chapitre 2 : Fondements théoriques et modélisation du système d'assistance à la mobilité

2.1. Introduction

L'étude des systèmes d'assistance à la mobilité implique une compréhension approfondie des mécanismes par lesquels l'information environnementale est perçue, structurée et exploitée. Au-delà de la simple collecte de données, il est essentiel de considérer la manière dont ces informations sont traitées, intégrées et restituées de façon cohérente avec les capacités perceptives et cognitives des utilisateurs.

La modélisation offre un cadre rigoureux pour représenter ces processus, en formalisant les interactions entre perception, traitement et action, ainsi que les contraintes temporelles et fonctionnelles qui régissent le fonctionnement des systèmes dans des environnements complexes et dynamiques. Elle permet également de clarifier les relations entre les différentes composantes et d'anticiper les comportements du dispositif face à des situations variées.

Ce chapitre se concentre sur les fondements théoriques et les approches de modélisation qui sous-tendent l'analyse des systèmes d'assistance à la mobilité, fournissant un cadre conceptuel structuré pour la réflexion, l'évaluation et la compréhension des dispositifs étudiés dans cette thèse.

2.2. Définition d'un système d'assistance à la mobilité

Un système d'assistance à la mobilité pour personnes déficientes visuelles est un dispositif technologique intégré visant à traduire les informations spatiales et dynamiques de l'environnement en représentations exploitables par l'utilisateur. Il a pour objectif de réduire l'incertitude liée à l'absence de perception visuelle, en fournissant une information continue, hiérarchisée et adaptée aux capacités sensorielles et cognitives.

Ce type de système intègre des fonctions essentielles de perception de l'environnement, de localisation de l'utilisateur, d'évaluation et d'évitement des obstacles, ainsi que de restitution par des canaux non visuels. Son efficacité repose sur la cohérence et l'intégration dynamique de ces fonctions, en tenant compte des contraintes de l'utilisateur et des caractéristiques des environnements intérieurs et extérieurs.

L'analyse de l'environnement, la typologie des obstacles et la définition des repères spatiaux constituent des bases théoriques fondamentales, permettant de générer une représentation exploitable de l'espace et d'anticiper les interactions avec les éléments statiques ou mobiles. Cette définition fournit un cadre conceptuel pour la modélisation fonctionnelle du système et pour la représentation de l'environnement, éléments centraux à la conception et à l'optimisation des dispositifs d'assistance [61].

La figure 2.1 illustre une vue fonctionnelle globale du système d'assistance pour personnes non-voyantes

.

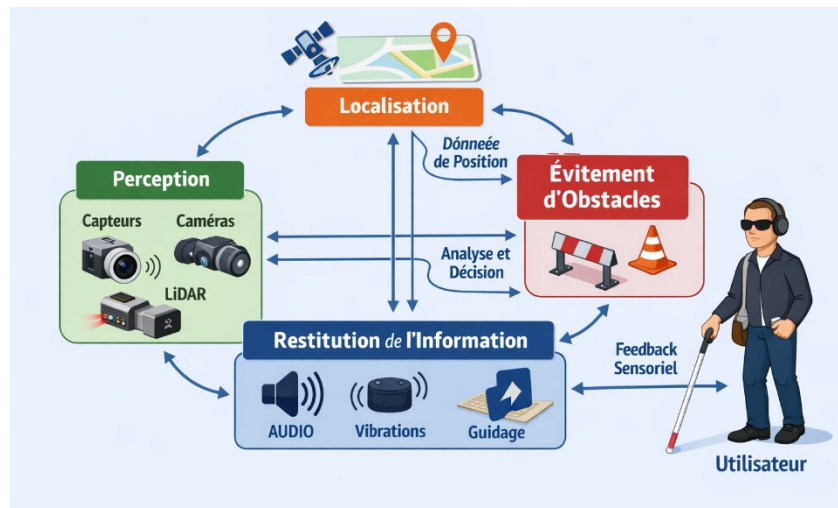


Figure 2.1. Schéma général d'un système d'assistance à la mobilité pour les personnes non-voyantes

2.3. Représentation de l'environnement pour la mobilité non-visuelle

La représentation de l'environnement est un élément central des systèmes d'assistance à la mobilité pour les personnes déficientes visuelles. Elle permet de structurer l'information spatiale en intégrant les objets, obstacles, trajectoires et relations spatiales, afin de transformer un environnement complexe en un modèle exploitable pour la localisation, la planification, l'évitement et la restitution. Ces modèles prennent en compte la topologie des lieux, les caractéristiques et la dynamique des obstacles, ainsi que la hiérarchisation de l'information nécessaire à la navigation, tout en intégrant l'incertitude inhérente à la perception [62-63].

Différentes approches, topologiques, métriques, hybrides ou sémantiques, offrent des compromis entre précision, lisibilité et capacité à représenter la dynamique de l'environnement. Elles fournissent une base conceptuelle pour formaliser le système, guider la planification des trajectoires et préparer les algorithmes de navigation, indépendamment des choix technologiques spécifiques.

2.3.1. Composants essentiels pour la modélisation de l'environnement non-visuel

La modélisation de l'environnement pour la navigation non-visuelle repose sur l'identification et la structuration des composants essentiels qui influencent la sécurité et l'orientation de l'utilisateur. Chaque élément est caractérisé par ses propriétés principales et ses relations avec les autres entités spatiales, formant une base exploitable pour la planification des trajectoires et la prise de décision. Parmi ces composants, on distingue : les obstacles fixes ou mobiles à hauteur de corps ou de visage, les obstacles au sol tels que marches, bordures ou trous, les repères directionnels et points de référence, ainsi que les éléments de signalisation fournissant des informations contextuelles et orientantes.

La formalisation de ces composants mobilise plusieurs approches complémentaires. La modélisation par graphes [64] traduit les relations spatiales en structures abstraites, facilitant la planification et la hiérarchisation des trajectoires. Les cartes sémantiques [65] enrichissent cette représentation en associant à chaque entité des propriétés interprétables, telles que le type, la dynamique et la criticité pour la sécurité, permettant au système de raisonner sur des situations complexes et d'émettre des alertes adaptées. La modélisation topologique [66] se concentre sur les relations entre entités plutôt que sur les coordonnées absolues, garantissant la continuité et la cohérence des chemins, ce qui est particulièrement utile dans les environnements intérieurs.

Pour gérer l'incertitude liée aux obstacles mobiles et aux mesures bruitées, des approches probabilistes et des réseaux de contraintes [67] sont employés. Les positions et mouvements des objets sont représentés par des distributions de probabilité, permettant de prédire leur évolution et de moduler la prise de décision en temps réel. Les réseaux de contraintes combinent les relations spatiales et les règles de navigation, assurant la sécurité et la faisabilité des trajectoires.

L'intégration de ces méthodes produit une représentation multi-échelle de l'environnement, capable de gérer simultanément les informations locales immédiates et la structure globale de l'espace, tout en anticipant la dynamique des obstacles. Cette structuration des composants essentiels constitue un fondement théorique central pour les systèmes d'assistance à la mobilité, orientant la collecte de données, la planification des trajectoires et la restitution d'informations, et garantissant une interaction autonome et sécurisée avec l'utilisateur.

2.3.2. Typologie des environnements et composantes spatiales

La compréhension des environnements traversés constitue un fondement central pour tout système d'assistance à la mobilité destiné aux personnes déficientes visuelles. Ces environnements se composent d'éléments dont la nature, la disposition et la dynamique influencent directement la sécurité et l'orientation de l'utilisateur. On distingue principalement deux grandes catégories : les environnements intérieurs et les environnements extérieurs, chacun présentant des composantes spécifiques et des interactions spatiales particulières qui conditionnent la navigation non-visuelle. La figure 2.2 illustre les composants essentiels pour la mobilité dans ces différents contextes



Figure 2.2. Typologie des environnements et composantes spatiales

- Les environnements intérieurs comprennent les bâtiments, couloirs, halls, escaliers, ascenseurs et autres espaces clos accessibles au public. Leurs composantes principales englobent des obstacles ponctuels et structuraux (colonnes, murs, portes, mobilier) ainsi que des changements de niveau (marches, rampes), qui définissent les trajectoires accessibles et imposent des contraintes de franchissement et de contournement. Les repères spatiaux, tels que les numéros de porte, les panneaux directionnels ou les intersections de corridors, fournissent des références essentielles pour la construction de cartes mentales et la planification des déplacements. La densité et la disposition de ces composants influencent la complexité de la navigation et la nécessité de stratégies d'évitement adaptées [68].
- Les environnements extérieurs regroupent les rues, trottoirs, places et espaces ouverts urbains ou semi-urbains. Les composantes critiques incluent des obstacles fixes (poteaux, bancs, bordures, arbres) et obstacles mobiles (véhicules, cyclistes, piétons), ainsi que les éléments de signalisation (passages piétons, feux, panneaux, marquages au sol). Les variations topographiques et les conditions environnementales (pluie, vent, bruit) ajoutent des contraintes supplémentaires, tandis que les repères auditifs naturels ou artificiels facilitent l'orientation [69].

Ces composantes peuvent être décrites par leurs propriétés opérationnelles : position, dimensions, type, dynamique et criticité. Les relations entre objets, proximité, alignement, accessibilité et continuité des trajectoires, sont centrales pour formaliser la navigation. Des approches telles que la modélisation par graphes, les cartes topologiques ou sémantiques offrent une vision structurée des parcours possibles et des interactions entre composants fixes et mobiles.

La catégorisation des composants critiques en fonction de leur rôle dans l'espace constitue un fondement théorique pour orienter le raisonnement du système, guider la planification des déplacements et hiérarchiser l'information pertinente. Elle fournit également un cadre pour déterminer les choix méthodologiques relatifs aux capteurs, à la modélisation et aux modes de restitution dans les systèmes d'assistance à la mobilité.

2.3.3. Propriétés critiques des obstacles et dynamique environnementale

La modélisation des obstacles et de la dynamique de l'environnement constitue un volet central des systèmes d'assistance à la mobilité pour les personnes déficientes visuelles. La sécurité et la planification des déplacements dépendent de la capacité du système à représenter les objets et leurs interactions dans l'espace, ainsi qu'à anticiper leur évolution dans le temps. Chaque obstacle peut être caractérisé par ses propriétés physiques (forme, dimensions, position), fonctionnelles (mobilité, criticité) et contextuelles (densité, influence sur la trajectoire). Ces informations permettent de hiérarchiser les éléments selon leur importance pour la navigation et de guider la prise de décision algorithmique.

Les obstacles fixes, tels que murs, colonnes ou mobilier, définissent les limites spatiales et servent de repères stables pour l'orientation et la planification. Les obstacles mobiles, comme piétons, véhicules ou cyclistes, introduisent une incertitude temporelle qui requiert des modèles prédictifs capables d'estimer leur position et leur trajectoire future. Les approches probabilistes, incluant les cartes d'occupation dynamiques ou les filtres de type Kalman, fournissent un cadre robuste pour traiter ces incertitudes et ajuster en temps réel les recommandations de trajectoire.

La dynamique environnementale englobe également les variations structurelles et contextuelles, qu'il s'agisse de modifications ponctuelles dans les espaces intérieurs ou des conditions extérieures telles que le trafic, le flux piétonnier, la topographie ou les conditions climatiques. Les modèles topologiques évolutifs et les représentations multi-échelles permettent de maintenir la cohérence des parcours et de représenter simultanément les informations locales et globales, offrant une vision intégrée de l'environnement.

L'intégration de ces propriétés dans la modélisation d'un système d'assistance établit un cadre théorique solide pour la conception des fonctions de détection, d'évitement, de planification et de restitution. Elle garantit que le dispositif peut anticiper les contraintes, hiérarchiser les informations pertinentes et soutenir une navigation sécurisée et efficace, tout en restant adaptable aux variations complexes et dynamiques des environnements traversés.

2.4. Besoins des personnes non-voyantes dans un système d'assistance à la mobilité

La conception d'un système d'assistance à la mobilité ne peut se limiter à des considérations technologiques ou algorithmiques : elle doit avant tout répondre aux besoins réels des utilisateurs. Ces besoins se fondent sur les contraintes rencontrées au quotidien par les personnes non-voyantes, leurs stratégies d'adaptation et les limites des aides existantes [70-71].

La mobilité sans vision repose sur une interaction continue entre perception non-visuelle, action motrice et compréhension de l'environnement. Dans ce cadre, l'assistance technologique n'est pas un substitut aux capacités de l'utilisateur, mais un support fonctionnel destiné à sécuriser le déplacement, réduire l'incertitude et renforcer l'autonomie. L'identification précise de ces besoins constitue un fondement essentiel pour orienter la conception du système, hiérarchiser ses fonctions et limiter la surcharge informationnelle ou cognitive.

Afin de traduire ces besoins en exigences exploitables pour la modélisation d'un système d'assistance, cette section adopte une structuration progressive. Les besoins sont analysés selon une logique fonctionnelle, allant des conditions minimales de sécurité vers des dimensions plus intégratives liées à l'orientation, au traitement de l'information et à l'usage durable du système.

2.4.1. Sécurité du déplacement et prévention des situations à risque

La sécurité constitue la priorité absolue pour les personnes non-voyantes lors de leurs déplacements. L'absence de vision expose en permanence à des risques liés aux obstacles fixes ou aux éléments dynamiques de l'environnement. La crainte de collision, de chute ou de désorientation influence directement le comportement de déplacement et peut limiter l'autonomie.

- Un besoin fondamental est la détection anticipée des obstacles. Les aides traditionnelles, comme la canne blanche, explorent uniquement l'espace proche et ne permettent pas d'identifier certains dangers, tels que les obstacles surélevés ou mobiles. Les utilisateurs ont besoin d'être informés suffisamment tôt pour adapter leur vitesse, leur trajectoire et leur posture, garantissant ainsi une navigation sécurisée.
- La compréhension du niveau de risque associé à chaque obstacle constitue un autre aspect crucial. Tous les obstacles ne nécessitent pas la même réaction : un obstacle au sol peut demander un contournement précis, tandis qu'un obstacle mobile exige une adaptation dynamique. Les utilisateurs ont besoin d'informations claires, exploitables immédiatement, qui ne génèrent ni ambiguïté ni surcharge cognitive.

- La stabilité lors du déplacement représente également un enjeu majeur. Les irrégularités du sol, les changements de niveau et les ruptures de continuité sont des sources fréquentes de chutes. L'assistance doit fournir des alertes fiables et cohérentes, permettant à l'utilisateur d'anticiper les mouvements et d'adopter une posture sécurisée.
- La prévention des risques repose sur la continuité et la fiabilité de l'information. Les interruptions ou signaux incohérents peuvent générer de l'anxiété et diminuer la confiance dans le dispositif. Les utilisateurs expriment un besoin de constance et de cohérence dans l'assistance, conditions essentielles pour l'acceptation et l'usage prolongé du système.
- Enfin, la sécurité est étroitement liée au sentiment de contrôle. L'assistance doit renforcer la capacité de l'utilisateur à décider de ses déplacements, plutôt que d'imposer des instructions rigides. La prévention des risques doit donc soutenir l'action tout en laissant à la personne non-voyante la maîtrise finale de ses choix.

Une fois la sécurité immédiate assurée, le déplacement ne se limite plus à l'évitement des dangers. Il implique la capacité à se situer, à maintenir une direction cohérente et à comprendre l'organisation de l'espace traversé, ce qui conduit aux besoins d'orientation.

2.4.2. Orientation spatiale et compréhension progressive de l'environnement

Au-delà de la sécurité immédiate, la mobilité autonome implique la capacité à se situer dans l'espace et à comprendre l'organisation de l'environnement traversé. Pour une personne non-voyante, l'orientation ne repose pas sur une perception globale et instantanée de l'espace, mais sur une construction progressive fondée sur des repères partiels, séquentiels et souvent multisensoriels. Cette particularité engendre un besoin fondamental d'assistance dans la structuration spatiale du déplacement.

- L'un des premiers besoins exprimés concerne la capacité à identifier des repères stables permettant de maintenir une direction cohérente. En l'absence de vision, la perte d'alignement ou la dérive progressive de la trajectoire est fréquente, notamment dans les espaces ouverts ou peu structurés. Les personnes non-voyantes ont besoin d'éléments de référence fiables leur permettant de conserver une orientation relative, que ce soit dans un environnement familier ou inconnu. La compréhension de l'espace ne se limite pas à la direction immédiate. Elle inclut également la capacité à appréhender la structure globale du lieu, à savoir la succession des zones traversées, les points de transition et les choix possibles de déplacement. Cette représentation mentale, souvent qualifiée de carte cognitive non visuelle, est construite à partir d'indices auditifs, tactiles ou proprioceptifs. Un système d'assistance est attendu pour soutenir cette construction, en fournissant des informations suffisamment structurées pour être intégrées mentalement, sans interrompre le flux naturel du déplacement.

- Un autre besoin essentiel concerne la distinction entre environnements familiers et non familiers. Dans un espace connu, la personne non-voyante s'appuie sur des routines et des repères mémorisés, tandis que dans un environnement nouveau, l'incertitude augmente fortement. L'assistance doit alors contribuer à réduire cette incertitude en aidant l'utilisateur à comprendre la logique de l'espace, sans imposer une description exhaustive qui serait difficilement exploitable en situation de mobilité.
- L'orientation implique également la capacité à anticiper les transitions spatiales, telles que l'entrée dans un bâtiment, le franchissement d'un carrefour ou le changement de niveau. Ces moments sont critiques, car ils nécessitent une adaptation rapide du comportement de déplacement. Les personnes non-voyantes expriment le besoin d'être préparées à ces transitions, afin de pouvoir ajuster leur posture, leur vitesse ou leur stratégie d'exploration.

Ces processus d'orientation reposent sur une interprétation continue de l'information disponible. Leur efficacité dépend donc directement de la manière dont cette information est sélectionnée, structurée et restituée à l'utilisateur, ce qui met en évidence les exigences cognitives associées.

2.4.3. Gestion de la charge cognitive et contrôle de l'information

La mobilité sans vision sollicite fortement les ressources cognitives de l'utilisateur, dont l'attention est mobilisée pour interpréter les signaux disponibles, comprendre l'environnement et anticiper les situations à risque. Dans ce contexte, la maîtrise de l'information délivrée par un système d'assistance constitue un besoin central.

Un système fournissant des messages trop nombreux ou excessivement détaillés peut rapidement devenir contre-productif. Les utilisateurs rencontrent des difficultés à traiter simultanément plusieurs informations, en particulier dans des environnements complexes ou bruyants. Le besoin principal réside donc dans la fourniture d'informations pertinentes, hiérarchisées et directement exploitables pour l'action.

La temporalité de la restitution joue un rôle déterminant. Une information transmise trop tôt peut perdre sa valeur opérationnelle, tandis qu'une information tardive peut compromettre la sécurité. L'assistance doit ainsi être synchronisée avec le rythme de déplacement et les variations de l'environnement, afin de s'intégrer de manière fluide dans l'action en cours.

La cohérence et la stabilité des messages sont également essentielles. Des informations contradictoires ou instables augmentent l'effort cognitif et diminuent la confiance dans le système. La gestion de l'information doit enfin tenir compte de la diversité des profils utilisateurs, en adaptant la quantité, la forme et le rythme de restitution.

Ainsi, la maîtrise de la charge cognitive constitue un critère fondamental pour la modélisation des systèmes d'assistance à la mobilité, conditionnant leur efficacité et leur acceptabilité par les utilisateurs, mais également la relation de confiance que l'utilisateur développe avec le dispositif, aspect déterminant pour son adoption à long terme.

2.4.4. Confiance, acceptabilité et continuité d'usage

Au-delà de la performance fonctionnelle, l'efficacité réelle d'un système d'assistance à la mobilité dépend largement de son acceptabilité par les personnes non-voyantes. Un dispositif, même techniquement performant, peut être peu utilisé ou abandonné s'il ne répond pas aux exigences liées à l'usage quotidien, au confort et au sentiment de confiance.

- La confiance dans le système constitue un besoin central. Elle se construit progressivement à partir de la fiabilité des informations fournies et de la cohérence du comportement du dispositif. Les utilisateurs ont besoin de percevoir une certaine prévisibilité dans les réactions du système afin de pouvoir anticiper son fonctionnement et l'intégrer dans leurs stratégies personnelles de déplacement. Un système perçu comme instable ou imprévisible tend à réduire la confiance et l'engagement à long terme.
- L'acceptabilité est également étroitement liée à la discrétion et à l'intégration sociale du dispositif. Les personnes non-voyantes privilégient des solutions qui s'intègrent naturellement dans leur quotidien, sans générer de stigmatisation. Le confort d'usage, la simplicité de prise en main et la compatibilité avec les aides existantes constituent des facteurs déterminants pour une adoption durable.
- L'usage quotidien implique par ailleurs des contraintes pratiques, telles que l'autonomie énergétique, la robustesse et la limitation des réglages ou interventions nécessaires. Un système d'assistance est attendu pour fonctionner de manière fiable dans des contextes variés, sans exiger une attention constante ou une maintenance complexe de la part de l'utilisateur.

Enfin, la confiance est indissociable du maintien de l'autonomie décisionnelle. L'assistance est perçue positivement lorsqu'elle soutient l'action sans s'y substituer, laissant à l'utilisateur la maîtrise finale de ses choix. La possibilité d'ignorer une information ou d'adapter son comportement demeure essentielle pour préserver le sentiment de contrôle.

Ces éléments concluent l'analyse des besoins en montrant que la mobilité assistée repose sur un enchaînement cohérent de fonctions, allant de la sécurité immédiate à l'intégration durable du système dans l'expérience quotidienne du déplacement sans vision.

2.5. Modélisation fonctionnelle du système

La modélisation fonctionnelle consiste à formaliser un système d'assistance à la mobilité comme un ensemble structuré de fonctions coopérantes, organisées selon des flux d'information clairement définis. Cette formalisation vise à décrire le fonctionnement global du système à un niveau abstrait, en identifiant les rôles respectifs des différentes fonctions et leurs interactions, sans référence à des choix technologiques ou implémentations particulières.

Une telle modélisation permet de raisonner sur la cohérence interne du système, la séquence des traitements et la circulation de l'information entre les différents composants fonctionnels. Elle constitue un outil conceptuel essentiel pour analyser le comportement du dispositif, évaluer sa capacité à répondre aux contraintes de fonctionnement et structurer les développements ultérieurs de manière rigoureuse.

En fournissant une représentation claire et indépendante de l'implémentation, la modélisation fonctionnelle établit un cadre de référence pour l'analyse, la comparaison et l'évolution des architectures de systèmes d'assistance, tout en garantissant la maîtrise de leur complexité.

2.5.1. Perception de l'environnement

La perception de l'environnement constitue une fonction fondamentale dans la modélisation d'un système d'assistance à la mobilité destiné aux personnes présentant une déficience visuelle. Elle vise à établir une description opérationnelle de l'espace de navigation en identifiant et en structurant les éléments pertinents pour le déplacement, servant ainsi de base aux fonctions d'analyse, de décision et de restitution.

Sur le plan fonctionnel, la perception permet d'extraire les informations essentielles relatives à l'environnement, de caractériser la configuration spatiale des obstacles et de distinguer les éléments stables des éléments évolutifs. La représentation produite est volontairement sélective, privilégiant les éléments critiques pour la sécurité et la continuité du déplacement plutôt qu'une description exhaustive de la scène. Cette formalisation constitue un cadre conceptuel nécessaire pour assurer la cohérence et la fiabilité du fonctionnement global du système. Plusieurs approches théoriques sont mobilisées pour formaliser cette fonction de perception et garantir sa robustesse dans des environnements variés :

- La vision par ordinateur repose sur l'acquisition et l'analyse automatiques d'images ou de séquences vidéo afin d'extraire des informations sémantiques sur la scène observée [72]. L'image est représentée sous forme matricielle et traitée à l'aide d'opérations mathématiques telles que le filtrage, la détection de contours, la segmentation et l'extraction de caractéristiques visuelles. Ces traitements permettent d'identifier les objets présents, d'en estimer la position relative et d'en suivre l'évolution temporelle. Dans le contexte de l'assistance à la mobilité, la vision par ordinateur contribue à la reconnaissance des obstacles et à la compréhension de la

structure de l'environnement, fournissant ainsi une représentation sémantique utile à la navigation. La Figure 2.3 illustre le principe général de fonctionnement de cette approche dans le cadre de la détection d'obstacles.

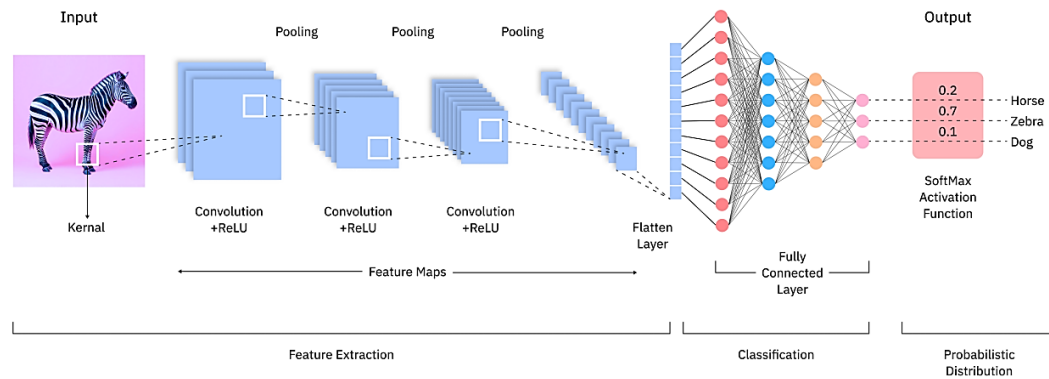


Figure 2.3. Principe général de fonctionnement de la vision par ordinateur

- Les capteurs de distance, tels que les capteurs ultrasonores, infrarouges, laser ou à temps de vol (Time-of-Flight, ToF), permettent de mesurer la distance séparant l'utilisateur des objets environnants [73-74]. Leur principe repose sur l'émission d'un signal physique vers l'environnement, suivi de la mesure du signal réfléchi afin d'estimer la distance par calcul du temps de propagation ou de la variation de phase. Ces capteurs fournissent des informations métriques directes, essentielles pour évaluer la proximité des obstacles et pour garantir un déplacement sécurisé, notamment dans les situations où les conditions visuelles sont défavorables ou insuffisantes. La figure 2.4 illustre le principe de fonctionnement des capteurs ToF.

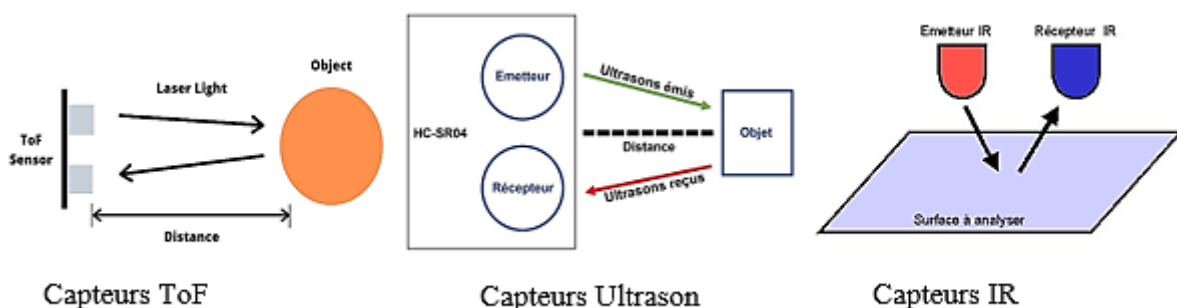


Figure 2.4. Principe de fonctionnement des capteurs de distance

- La fusion multi-capteurs constitue une approche centrale pour renforcer la fiabilité et la continuité de la perception [75]. Elle consiste à combiner les informations issues de différentes sources sensorielles afin de pallier les limitations propres à chaque capteur pris individuellement. Les données sont intégrées à l'aide de modèles mathématiques ou probabilistes permettant de réduire l'incertitude, de corriger les erreurs de mesure et d'assurer une perception cohérente de l'environnement. Par exemple, les informations sémantiques fournies par la vision par ordinateur peuvent être associées aux mesures métriques précises issues des capteurs de distance. Cette complémentarité permet d'obtenir une perception robuste, adaptée aux environnements complexes, dynamiques ou soumis à des perturbations telles que la faible luminosité, le bruit ou les variations de texture. La figure 2.5 illustre le principe de fonctionnement de la fusion multi-capteurs.

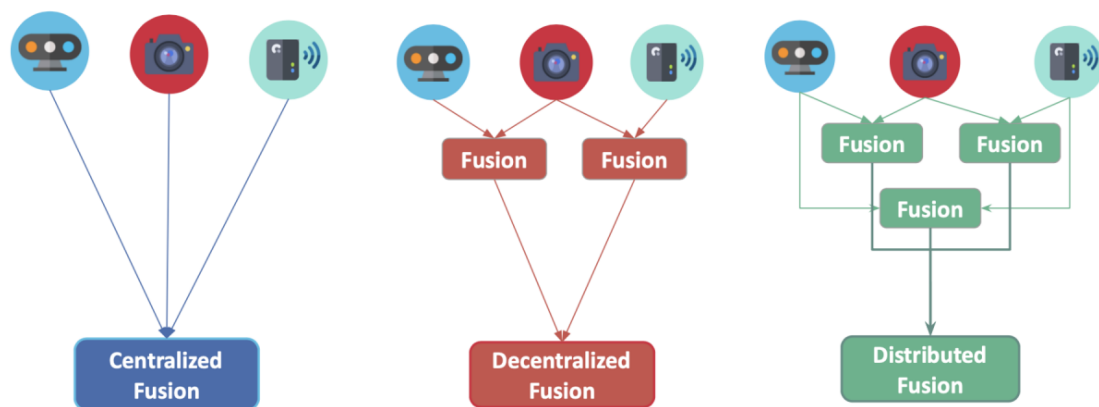


Figure 2.5. Principe de fonctionnement de la fusion multicapteurs

La fonction de perception ne se limite pas à une problématique technologique, mais s'inscrit dans une logique d'assistance centrée sur l'utilisateur. Elle vise à fournir des informations fiables, pertinentes et directement exploitables pour soutenir les capacités de déplacement autonome, réduire le risque d'accident et améliorer la confiance du non-voyant dans le système. Cette perception constitue ainsi la base fonctionnelle sur laquelle s'appuient les mécanismes de localisation, d'évitement d'obstacles et de restitution de l'information, assurant une interaction fluide et sécurisée entre l'utilisateur et son environnement.

2.5.2. Localisation de l'utilisateur

La localisation de l'utilisateur constitue une fonction clé d'un système d'assistance à la mobilité pour les personnes présentant une déficience visuelle. Elle vise à estimer de manière continue la position de l'utilisateur dans l'espace, ainsi que son orientation et l'évolution de son déplacement, afin de fournir un cadre de référence stable pour l'interprétation de l'environnement et la prise de décision. Du point de vue du non-voyant, cette fonction permet de situer son corps par rapport aux obstacles, aux repères

spatiaux et aux trajectoires accessibles, conditionnant ainsi la cohérence et la fiabilité de l'assistance fournie [76].

Contrairement à la perception, qui renseigne sur l'état de l'environnement, la localisation apporte une information centrée sur l'utilisateur lui-même. Elle permet d'ancrer les éléments perçus dans un référentiel spatial cohérent et de suivre l'évolution du déplacement dans le temps. Cette capacité est essentielle pour assurer la continuité de la navigation, éviter les désorientations et adapter les recommandations du système aux mouvements réels de l'utilisateur, y compris dans des environnements partiellement connus, dynamiques ou contraints.

La localisation repose sur l'exploitation conjointe de plusieurs sources d'information, combinant des mesures absolues et relatives afin de compenser les limitations propres à chaque modalité de mesure et d'assurer une estimation robuste et continue de la position :

- Les systèmes de positionnement par satellite, tels que le GPS ou plus largement les systèmes GNSS, fournissent une estimation absolue de la position géographique de l'utilisateur à partir de signaux émis par des constellations de satellites [77]. Basés sur le principe de la trilatération, ces systèmes permettent de situer l'utilisateur dans un référentiel global. Dans le contexte de la mobilité assistée, ils constituent une référence spatiale essentielle pour les déplacements en environnements extérieurs ouverts, notamment pour le guidage sur de longues distances et la contextualisation des informations de navigation. La figure 2.6 montre le principe de la trilatération.

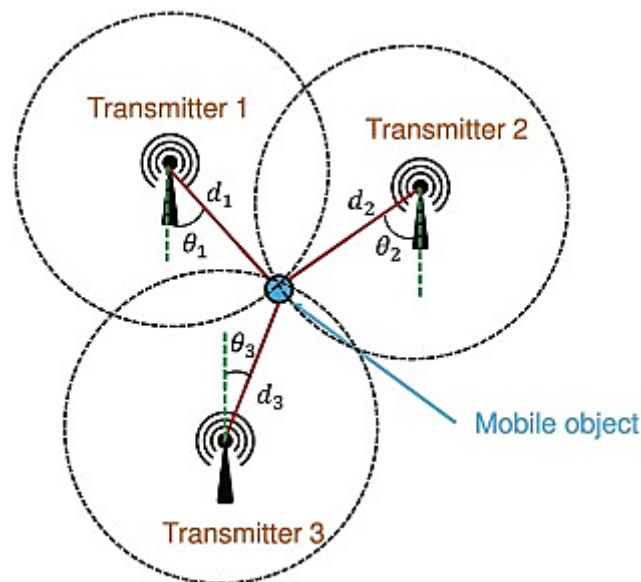


Figure 2.6. Le principe de la trilatération.

- Les unités de mesure inertielle (IMU) permettent de mesurer les accélérations et les vitesses angulaires du corps de l'utilisateur selon plusieurs axes [78]. Par intégration temporelle de ces

mesures, il devient possible d'estimer les déplacements relatifs et l'orientation, indépendamment de toute infrastructure externe. Cette approche est particulièrement pertinente dans les environnements intérieurs ou couverts, où les signaux satellitaires sont dégradés ou absents, et permet d'assurer un suivi continu du mouvement, essentiel à la fluidité de la navigation. La figure 2.7 représente un modèle IMU précisant les paramètres essentiels pour la localisation.

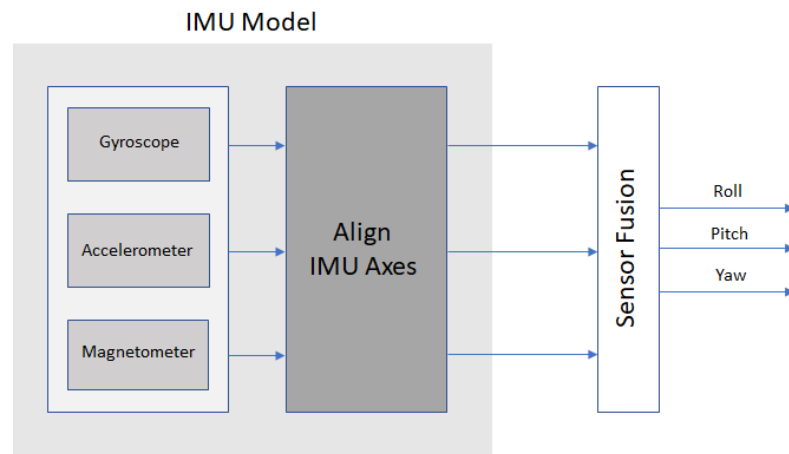


Figure 2.7. Le modèle IMU pour la localisation

- Les méthodes d'estimation probabiliste, telles que les filtres de Kalman et leurs extensions non linéaires, jouent un rôle central dans la gestion de l'incertitude associée aux mesures de localisation [79-80]. Ces approches reposent sur la prédiction de l'état futur de l'utilisateur à partir d'un modèle de mouvement, suivie d'une correction basée sur les observations réelles fournies par les capteurs. Elles permettent d'obtenir une estimation lissée et cohérente de la position et de l'orientation, tout en tenant compte du bruit, des erreurs de mesure et des perturbations environnementales.
- La fusion multi-capteurs constitue enfin le principe fondamental assurant la robustesse de la localisation. En combinant les informations issues des systèmes satellitaires, des capteurs inertiels et, le cas échéant, d'autres sources de mesure, le système exploite la complémentarité entre précision globale, continuité temporelle et sensibilité locale. Cette intégration permet de maintenir une estimation fiable de la position de l'utilisateur dans des contextes variés, qu'il s'agisse d'environnements intérieurs, extérieurs ou mixtes.

La fonction de localisation ne vise pas uniquement la précision géométrique, mais la cohérence perceptive et fonctionnelle de l'assistance fournie. Elle garantit que les informations restituées au non-voyant sont spatialement pertinentes, synchronisées avec ses mouvements et adaptées à la situation réelle de déplacement. La localisation constitue ainsi un fondement indispensable pour les fonctions

d'évitement d'obstacles, de guidage et de restitution de l'information, contribuant directement à la sécurité, à l'autonomie et à la confiance de l'utilisateur dans le système d'assistance.

2.5.3. Évitement d'obstacles

L'évitement d'obstacles constitue une fonction centrale du système d'assistance à la mobilité, directement liée à la sécurité physique et à l'autonomie de déplacement des personnes présentant une déficience visuelle. Cette fonction vise à interpréter les informations issues de la perception de l'environnement et de la localisation de l'utilisateur afin de prévenir les collisions et de guider le déplacement de manière fluide et adaptée [81]. Du point de vue de l'utilisateur non-voyant, l'évitement d'obstacles permet de transformer une perception indirecte et partielle de l'environnement en actions de déplacement cohérentes, sécurisées et anticipatives.

La complexité de cette fonction réside dans la nature hétérogène et dynamique de l'environnement de déplacement. Le système doit prendre en compte simultanément des obstacles statiques, tels que les poteaux, murs ou marches, et des obstacles dynamiques, comme les piétons ou objets en mouvement, dont la trajectoire et la vitesse peuvent évoluer de manière imprévisible. L'évitement repose ainsi sur des mécanismes capables de gérer l'incertitude, le bruit des capteurs et les variations rapides de la scène, tout en assurant une réactivité compatible avec un déplacement naturel.

- Une première famille d'approches repose sur la logique floue, particulièrement adaptée aux contextes réels caractérisés par des informations imprécises ou incomplètes [82]. Cette approche permet de représenter la proximité, la vitesse ou le niveau de risque des obstacles sous forme de degrés continus plutôt que de décisions binaires. Les règles floues traduisent ces degrés en actions graduelles, telles que l'ajustement progressif de la direction, la modulation de la vitesse de déplacement ou l'intensification des signaux de restitution. Cette continuité décisionnelle est essentielle pour garantir un comportement du système compatible avec les capacités perceptives et motrices de l'utilisateur. La figure 2.8 illustre un système général d'inférence d'un contrôleur flou.

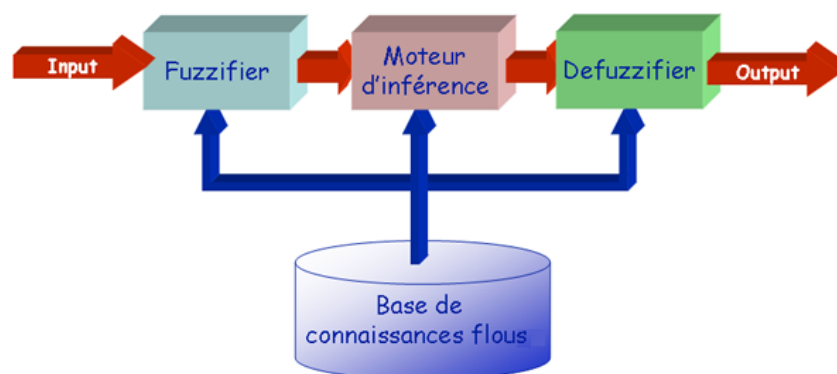


Figure 2.8. Un système général d'inférence d'un contrôleur flou.

- L'évitement d'obstacles s'appuie également sur des méthodes de planification de trajectoire, qui consistent à évaluer les chemins possibles dans l'espace de navigation en tenant compte des contraintes de sécurité, de confort et de continuité du déplacement [83-84]. Les approches basées sur les graphes permettent d'identifier des trajectoires optimales dans des environnements structurés, tandis que les méthodes probabilistes explorent l'espace de navigation de manière plus flexible dans des contextes partiellement connus ou dynamiques. Ces approches offrent un cadre théorique pour anticiper les interactions entre l'utilisateur et les obstacles, et pour ajuster le déplacement avant qu'une situation critique ne survienne.

Dans les environnements évolutifs, le recours au contrôle optimal et à la programmation dynamique permet d'intégrer explicitement des critères multiples dans la prise de décision. La trajectoire n'est alors pas uniquement définie par la distance minimale, mais par une fonction de coût globale intégrant la sécurité, la fluidité du mouvement et le confort perceptif. Le contrôle prédictif basé sur un modèle permet d'actualiser en continu les décisions de déplacement en fonction de l'évolution de l'environnement et des actions de l'utilisateur, assurant ainsi une adaptation en temps réel.

La gestion des obstacles dynamiques nécessite enfin l'intégration de modèles prédictifs de mouvement, capables d'estimer l'évolution future des objets environnants. Ces modèles exploitent les informations temporelles issues de la perception pour anticiper les trajectoires probables des obstacles et ajuster le déplacement de l'utilisateur en conséquence. Les approches probabilistes permettent de représenter l'incertitude associée à ces prédictions, tandis que les modèles issus de l'apprentissage automatique peuvent capturer des comportements complexes et non linéaires observés dans des environnements réels.

L'intégration de ces différentes approches confère à la fonction d'évitement d'obstacles un caractère anticipatif plutôt que purement réactif. Le système ne se contente pas de signaler un danger imminent, mais accompagne le déplacement en proposant des ajustements continus et cohérents, en tenant compte de la position de l'utilisateur, de la dynamique de l'environnement et des contraintes de confort. Cette fonction assure ainsi la continuité entre perception, localisation et décision, et contribue directement à renforcer la confiance de l'utilisateur dans le système d'assistance, condition essentielle à une mobilité autonome et sécurisée.

2.5.4. Restitution de l'information

La restitution de l'information constitue l'interface directe entre le système et l'utilisateur, et représente un élément déterminant pour la sécurité et l'efficacité de la mobilité assistée [85]. Elle a pour objectif de traduire les données issues de la perception, de la localisation et de l'évitement d'obstacles en signaux compréhensibles, exploitables et adaptés aux capacités sensorielles et cognitives des personnes non-

voyantes. Une restitution efficace permet à l'utilisateur de percevoir l'environnement de manière synthétique et anticipative, tout en soutenant la prise de décision et le contrôle de la trajectoire.

- La restitution repose sur le codage multimodal, qui exploite différents canaux sensoriels pour transmettre des informations complémentaires. Le canal auditif fournit des indications directionnelles et des alertes de proximité, modulées en fonction de la distance, de la criticité de l'obstacle ou de la trajectoire optimale [86]. Le canal tactile, via des dispositifs vibrotactiles ou haptiques portés sur le corps ou la canne, informe sur la nature, la position et l'évolution des obstacles, ainsi que sur les changements de direction ou les points de repère [87]. Cette combinaison multimodale permet de pallier la déficience visuelle en tirant parti de la plasticité sensorielle et cognitive de l'utilisateur, tout en assurant une perception continue et intuitive de l'environnement.

Un aspect essentiel de la restitution est la hiérarchisation de l'information. Dans un environnement complexe, toutes les données ne peuvent pas être transmises simultanément sans risque de surcharge cognitive. Le système sélectionne et priorise les informations critiques, telles que la présence d'obstacles immédiats ou la direction à suivre, tout en relayant secondairement des informations moins urgentes. Cette hiérarchisation repose sur une modélisation de la perception et de l'attention humaine, et permet d'optimiser la réactivité et la sécurité sans perturber la fluidité de la mobilité.

- La restitution fonctionne également selon le principe des boucles de rétroaction continue, inspirées du contrôle en boucle fermée. Le système adapte en permanence les signaux délivrés en fonction des actions de l'utilisateur et de l'évolution de l'environnement [88]. Par exemple, la vitesse de déplacement, l'intensité ou la fréquence des vibrations tactiles, ou encore la modulation des signaux sonores peuvent être ajustées en temps réel si un obstacle apparaît soudainement ou si l'utilisateur change de direction. Cette adaptation dynamique assure une navigation fluide et sécurisée, renforçant la confiance de l'utilisateur et son autonomie.
- Enfin, la restitution de l'information intègre les principes d'ergonomie cognitive pour que les signaux transmis soient rapidement interprétables et aisément intégrables dans le schéma mental de navigation. Les informations sont calibrées pour réduire la charge mentale et faciliter l'anticipation des obstacles et des trajectoires [89]. Cette dimension est essentielle pour que l'utilisateur puisse réagir efficacement, planifier ses mouvements et maintenir un contrôle permanent sur son déplacement, même dans des environnements denses, complexes ou en évolution rapide.

En combinant ces principes, la restitution de l'information devient un vecteur d'interaction intelligent et adaptatif, transformant les données brutes issues des capteurs en directives opérationnelles et intuitives pour l'utilisateur. Elle assure l'intégration fluide des fonctions de perception, de localisation

et d'évitement, permettant une mobilité autonome, sécurisée et confortable pour les personnes présentant une déficience visuelle.

2.5.5. Interactions et flux d'information

L'efficacité d'un système d'assistance à la mobilité pour les personnes présentant une déficience visuelle repose sur l'intégration harmonieuse de ses fonctions principales et sur la fluidité des flux d'information entre les différentes composantes et l'utilisateur. Ces fonctions, perception, localisation, évitement d'obstacles et restitution de l'information, interagissent de manière continue, formant un système dynamique capable de répondre en temps réel aux variations de l'environnement et aux actions de l'utilisateur.

- La première étape de ce flux consiste en la perception, qui collecte et structure les informations relatives à l'environnement immédiat et lointain. Les données issues des capteurs, qu'elles proviennent de la vision par ordinateur, des capteurs de distance ou des dispositifs multi-capteurs, sont traitées et analysées pour identifier les obstacles, les repères et les structures spatiales. Ces informations constituent le point de départ du raisonnement du système et alimentent simultanément la localisation et l'évitement.
- La localisation de l'utilisateur constitue le second élément du flux. Les mesures absolues et relatives sont combinées par des algorithmes de fusion multi-capteurs et des méthodes probabilistes afin de produire une estimation continue et fiable de la position, de l'orientation et de la vitesse de déplacement. Cette localisation sert de référence spatiale pour la planification des trajectoires et l'ajustement dynamique du comportement du système face aux obstacles détectés.
- La fonction d'évitement d'obstacles agit comme un module décisionnel, intégrant les informations issues de la perception et de la localisation pour générer des trajectoires sûres et fluides. Les algorithmes utilisés, qu'ils soient basés sur la logique floue, la programmation dynamique, les graphes ou les modèles prédictifs de mouvement, évaluent en permanence la configuration de l'environnement, anticipent l'évolution des obstacles et ajustent la vitesse et la direction du déplacement. L'évitement n'est ainsi pas un simple mécanisme réactif, mais une action planifiée et adaptative, guidée par la criticité et la dynamique des éléments rencontrés.
- La restitution de l'information constitue l'interface finale de ce flux, traduisant les décisions du système en signaux perceptibles par l'utilisateur. L'information restituée est hiérarchisée, multimodale et modulable en fonction des actions de l'utilisateur et de la complexité de l'environnement. Les boucles de rétroaction continue assurent une adaptation dynamique,

permettant de corriger les trajectoires et de renforcer la confiance de l'utilisateur dans le dispositif.

Le flux d'information global se caractérise par une interaction constante entre les modules : la perception alimente la localisation et l'évitement, la localisation guide la planification des trajectoires, l'évitement détermine les ajustements nécessaires, et la restitution informe l'utilisateur pour lui permettre d'agir en conséquence. Cette architecture intégrée garantit une cohérence spatiale et temporelle des informations, minimise l'incertitude et réduit la charge cognitive sur l'utilisateur, tout en maximisant la sécurité et l'autonomie de déplacement.

Enfin, l'interaction utilisateur-système joue un rôle central : chaque action de l'utilisateur, qu'il s'agisse de changements de direction, de vitesse ou d'arrêt, est captée et répercutée dans le système, qui ajuste instantanément ses recommandations. Cette boucle adaptative crée un système interactif et prédictif, où la coordination entre perception, localisation, évitement et restitution permet de maintenir une mobilité autonome, sécurisée et confortable dans des environnements variés, intérieurs comme extérieurs, statiques comme dynamiques.

2.6. Principes ergonomiques et cognitifs pour l'assistance à la mobilité

La conception d'un dispositif d'assistance à la mobilité destiné aux personnes présentant une déficience visuelle doit reposer sur une approche multidimensionnelle, combinant les connaissances issues des sciences biomédicales, de l'ergonomie, de la psychologie cognitive et de l'ingénierie des systèmes. L'objectif central est de permettre un usage intuitif, sécurisé et confortable, tout en minimisant la charge cognitive et physique imposée à l'utilisateur. Cette démarche repose sur la compréhension des besoins sensoriels, cognitifs et comportementaux des personnes non-voyantes et sur la capacité à traduire ces besoins en contraintes et principes de conception techniques [90].

- Un des fondements essentiels réside dans la légèreté et la compacité du dispositif, qui conditionne directement le confort et la mobilité. Le système ne doit pas interférer avec la posture naturelle, l'équilibre ou la fluidité des mouvements. Les contraintes biomécaniques liées au port prolongé sont particulièrement critiques : un poids excessif ou une distribution inadaptée peut provoquer de la fatigue musculaire, altérer la stabilité et limiter l'autonomie de déplacement. Les dispositifs modulaires ou portables sont donc privilégiés, permettant un ajustement selon les préférences de l'utilisateur et la nature de l'environnement.
- La disposition des capteurs et des interfaces de restitution constitue un second pilier ergonomique. Les capteurs doivent être positionnés pour couvrir l'ensemble des obstacles potentiels à hauteur de tête, de tronc et au sol, tout en restant discrets et non intrusifs. L'objectif

est de garantir une perception complète de l'environnement sans gêner la mobilité ou attirer inutilement l'attention de l'utilisateur. Cette disposition doit également tenir compte des principes biomécaniques et sensoriels : les capteurs ne doivent pas restreindre le champ de mouvement, et les interfaces de retour doivent être accessibles sans nécessiter des gestes complexes ou une adaptation cognitive prolongée.

- L'optimisation des interfaces sensorielles est directement liée aux capacités perceptives des non-voyants. Les canaux auditifs et tactiles sont utilisés de manière complémentaire, exploitant la plasticité sensorielle pour transmettre efficacement les informations critiques. Les signaux tactiles, par exemple, peuvent être localisés sur différentes parties du corps pour transmettre la direction ou la proximité d'obstacles, tandis que les signaux auditifs peuvent coder la distance, la vitesse et la nature des objets. L'intégration de ces modalités doit être hiérarchisée : les informations essentielles à la sécurité et à la navigation doivent primer, tandis que les informations contextuelles moins urgentes peuvent être délivrées de manière différée ou subtile, évitant une surcharge cognitive.
- La réduction de la charge cognitive constitue un aspect fondamental de la conception biomédicale. L'utilisateur ne doit pas être obligé d'apprendre des séquences complexes pour interpréter les signaux, ni de mémoriser des cartes mentales trop détaillées. Les retours doivent être intuitifs et exploitables de manière quasi-automatique, permettant à l'utilisateur de maintenir son attention sur la navigation réelle et sur l'environnement immédiat. La hiérarchisation de l'information, la cohérence des signaux et la constance de leur codage jouent un rôle central pour assurer l'adoption rapide et sécurisée du système.
- Un autre principe clé est l'adaptabilité du système aux capacités et aux préférences individuelles. Les utilisateurs présentent des sensibilités auditives et tactiles différentes, ainsi que des expériences variées en matière de mobilité. La personnalisation des seuils d'alerte, de la fréquence ou de l'intensité des signaux, et la possibilité de sélectionner des modes de restitution spécifiques, garantissent que le dispositif est réellement accessible à tous, tout en optimisant le confort et la sécurité. Cette approche centrée sur l'utilisateur permet de renforcer la confiance dans le système et de favoriser l'autonomie.
- Enfin, la conception doit prendre en compte la sécurité physique et psychologique. La confiance de l'utilisateur dans le dispositif repose sur sa capacité à anticiper les obstacles et à fournir des informations fiables et cohérentes. Les principes cognitifs et ergonomiques sont donc étroitement liés à la fiabilité technique : les signaux doivent être suffisamment précis pour éviter des mouvements brusques ou dangereux, tout en restant intuitifs et faciles à interpréter.

L'interaction entre perception, restitution et action constitue ainsi un cycle dynamique, dans lequel le dispositif et l'utilisateur agissent en coordination continue pour garantir une navigation sécurisée et efficace.

En résumé, les principes ergonomiques et cognitifs établissent un cadre théorique pour le design du système, en combinant confort, sécurité, efficacité et adaptabilité. Ils servent de fondement pour la modélisation intégrée du système et orientent les choix techniques, depuis le positionnement des capteurs jusqu'à la conception des interfaces sensorielles et des stratégies d'apprentissage utilisateur.

2.7. Cadre conceptuel de conception de la plateforme d'assistance à la mobilité

La conception d'un système d'assistance à la mobilité destiné aux personnes non-voyantes ne peut se limiter à la définition de fonctions algorithmiques ou à l'optimisation des performances de détection et de navigation. Elle repose sur un cadre conceptuel plus large, intégrant simultanément des contraintes biomédicales, ergonomiques, cognitives et technologiques. Ce cadre vise à assurer que la plateforme d'assistance s'inscrive de manière harmonieuse dans le déplacement quotidien de l'utilisateur, sans perturber ses capacités sensorielles restantes, sa posture, ni ses stratégies naturelles d'exploration de l'environnement. Dans cette perspective, la plateforme d'implémentation constitue une composante centrale du système, car elle conditionne la manière dont les fonctions de perception, de localisation, d'évitement et de restitution sont concrètement déployées. Les choix relatifs à l'intégration corporelle, au positionnement des capteurs, aux capacités de calcul embarquées et à l'interaction utilisateur ont un impact direct sur la sécurité, la fiabilité et l'acceptabilité du dispositif [91]. Cette section établit ainsi un cadre théorique structuré pour la conception de la plateforme, en s'appuyant sur des principes issus de la biomécanique, de l'ingénierie des systèmes embarqués et des sciences cognitives appliquées à la mobilité assistée.

2.7.1. Contraintes biomécaniques et intégration corporelle du dispositif

L'intégration corporelle d'un système d'assistance à la mobilité constitue une problématique centrale en ingénierie biomédicale, dans la mesure où le dispositif interagit directement avec le corps humain en situation de déplacement. Le port du système ne doit ni altérer la posture naturelle, ni introduire des contraintes mécaniques susceptibles de générer une fatigue excessive, des compensations motrices ou des risques musculosquelettiques à long terme.

D'un point de vue biomécanique, la répartition de la masse du dispositif, sa stabilité lors de la marche et son interaction avec les segments corporels sont des paramètres déterminants. Un système trop lourd ou mal équilibré peut modifier la cinématique de la marche, perturber la coordination intersegmentaire

et réduire la fluidité du déplacement. La conception doit donc viser une masse minimale, une distribution homogène des composants et une fixation qui respecte les axes de mouvement naturels du corps.

L'intégration corporelle doit également préserver les canaux sensoriels déjà fortement sollicités chez les personnes non-voyantes, en particulier l'audition et le toucher. Le dispositif ne doit pas obstruer les oreilles, entraver l'utilisation de la canne blanche ou interférer avec les stratégies tactiles de reconnaissance de l'environnement. Cette contrainte impose une réflexion approfondie sur les zones du corps les plus appropriées pour l'implantation du système, telles que le torse, la ceinture ou certaines parties du tronc, qui offrent un compromis entre stabilité mécanique et neutralité fonctionnelle.

Enfin, l'intégration corporelle s'inscrit dans une logique de port prolongé. Le système doit pouvoir être utilisé sur de longues durées sans inconfort, sans provoquer de surcharge posturale et sans nécessiter d'ajustements fréquents. Ces exigences biomécaniques constituent une base fondamentale pour garantir que le dispositif s'intègre naturellement dans la mobilité quotidienne de l'utilisateur.

2.7.2. Architecture perceptive embarquée et positionnement des capteurs

L'architecture perceptive embarquée détermine la capacité du système à capter de manière fiable et pertinente les informations issues de l'environnement. Le positionnement des capteurs ne relève pas uniquement d'un choix technique, mais d'un compromis complexe entre couverture perceptive, robustesse des mesures, contraintes biomécaniques et cohérence fonctionnelle avec les autres modules du système.

Le champ de perception doit permettre la détection des obstacles situés dans la zone critique de déplacement, incluant les obstacles au sol, à hauteur du corps et dans l'espace de progression immédiat. Le positionnement des capteurs doit ainsi être pensé en relation directe avec la géométrie corporelle de l'utilisateur et la dynamique de la marche. Une implantation trop basse limite la perception des obstacles en hauteur, tandis qu'un positionnement trop élevé peut réduire la précision pour les obstacles proches du sol.

La stabilité des mesures constitue également un critère majeur. Les capteurs embarqués sont soumis aux vibrations, aux oscillations et aux accélérations induites par le mouvement du corps. Leur emplacement doit minimiser ces perturbations afin de garantir une acquisition fiable des données. Cette contrainte favorise des zones du corps présentant une stabilité relative lors de la marche, comme le tronc, par rapport aux membres supérieurs ou inférieurs.

Par ailleurs, l'architecture perceptive doit être conçue de manière modulaire, permettant l'intégration cohérente de capteurs complémentaires. Cette modularité facilite l'adaptation du système à différents contextes d'usage, tout en maintenant une architecture globale structurée et extensible. Le positionnement des capteurs devient alors un élément clé de la cohérence entre perception, localisation et évitement, sans introduire de redondance fonctionnelle inutile.

2.7.3. Exigences computationnelles et garanties de fonctionnement en temps réel

Le fonctionnement d'un système d'assistance à la mobilité repose sur le traitement continu de flux de données sensoriels et sur la génération de décisions en temps réel. Les exigences computationnelles associées à ces traitements imposent des contraintes fortes sur la plateforme d'implémentation, en particulier dans un contexte embarqué.

La latence constitue un paramètre critique. Les informations perceptives doivent être traitées et restituées dans des délais compatibles avec la dynamique de la marche humaine. Un retard excessif dans le traitement peut entraîner une inadéquation entre la situation réelle et l'information fournie à l'utilisateur, compromettant ainsi la sécurité du déplacement. La plateforme doit donc offrir des garanties de temps de réponse déterministes ou quasi déterministes.

Les ressources de calcul disponibles doivent être dimensionnées en fonction de la complexité des algorithmes, tout en respectant les contraintes de consommation énergétique. Un équilibre doit être trouvé entre la richesse des traitements réalisés et l'autonomie du système. Cette contrainte est particulièrement importante pour un dispositif destiné à un usage quotidien, où la fiabilité énergétique conditionne directement l'utilisabilité.

La plateforme doit également assurer une gestion robuste des défaillances partielles, telles que la perte temporaire de certaines données sensorielles ou des fluctuations de performance. La conception doit intégrer des mécanismes de tolérance aux fautes et de dégradation contrôlée du service, garantissant que le système reste fonctionnel même en conditions non idéales. Ces exigences computationnelles constituent un pilier fondamental de la fiabilité globale du dispositif.

2.7.4. Impact cognitif, apprentissage utilisateur et appropriation fonctionnelle

Au-delà des performances techniques, l'efficacité d'un système d'assistance à la mobilité dépend de sa capacité à être compris, appris et utilisé de manière intuitive par l'utilisateur. L'impact cognitif du système doit être soigneusement maîtrisé afin de ne pas surcharger les capacités attentionnelles déjà fortement sollicitées lors de la navigation non visuelle.

Le système doit fournir une information hiérarchisée, pertinente et contextualisée, évitant toute saturation sensorielle. La restitution doit soutenir la prise de décision sans imposer un effort d'interprétation excessif. Cette approche favorise une appropriation progressive du dispositif, où l'utilisateur développe une confiance dans le système et intègre son fonctionnement dans ses stratégies naturelles de déplacement.

L'apprentissage du système doit être rapide et ne pas nécessiter de phases prolongées d'entraînement. La logique de fonctionnement doit être cohérente et prévisible, permettant à l'utilisateur d'anticiper les

réactions du dispositif. Cette prévisibilité est essentielle pour instaurer une relation de confiance entre l'utilisateur et le système, condition indispensable à une utilisation régulière et sécurisée.

L'appropriation fonctionnelle s'inscrit ainsi dans une dynamique d'interaction homme-système, où le dispositif devient un prolongement cognitif de l'utilisateur, sans se substituer à ses capacités décisionnelles. Cette dimension constitue un enjeu central pour l'acceptabilité et la pérennité du système.

2.7.5. Rôle de la plateforme d'implémentation dans l'architecture globale du système

La plateforme d'implémentation ne constitue pas un simple support matériel des algorithmes, mais un élément structurant de l'architecture globale du système d'assistance à la mobilité. Elle conditionne l'intégration cohérente des fonctions définies au niveau théorique et leur traduction en un dispositif opérationnel.

Elle assure l'articulation entre les différentes couches fonctionnelles du système, depuis l'acquisition des données jusqu'à la restitution de l'information. Son architecture doit permettre une circulation fluide des données, une synchronisation efficace des modules et une gestion centralisée des ressources. Cette organisation garantit la cohérence globale du système et facilite son évolution future.

La plateforme joue également un rôle clé dans la validation expérimentale et la reproductibilité des résultats. Un cadre d'implémentation clairement défini permet de relier les performances observées aux choix architecturaux effectués, renforçant ainsi la rigueur scientifique de l'approche. Elle constitue enfin un point de transition naturel vers les chapitres suivants, dédiés à la conception détaillée, à l'implémentation et à l'évaluation du système.

2.8. Conclusion

L'assistance à la mobilité non visuelle s'appuie sur une compréhension intégrée de l'environnement, des capacités de l'utilisateur et des mécanismes décisionnels du système. La structuration des composantes spatiales, la caractérisation des obstacles et l'analyse de leur dynamique permettent de transformer un espace complexe et évolutif en informations exploitables pour la navigation autonome. Cette organisation constitue la base sur laquelle reposent les fonctions de perception, de localisation, d'évitement et de restitution de l'information. L'identification des besoins fondamentaux des personnes non-voyantes met en évidence l'importance d'une approche centrée sur la sécurité, l'orientation et la gestion de la charge cognitive. La modélisation fonctionnelle traduit ces besoins en mécanismes opérationnels cohérents, assurant une interaction continue entre l'utilisateur et son environnement. L'intégration des principes ergonomiques et cognitifs garantit que cette interaction demeure naturelle, fiable et compatible avec un usage quotidien.

L'ensemble de ces fondements théoriques établit une architecture conceptuelle cohérente, orientée vers une assistance à la mobilité à la fois robuste, intelligible et adaptée aux contraintes humaines. Cette base permet d'aborder la conception et l'implémentation du système avec une vision structurée, assurant la continuité entre les choix théoriques et leur traduction méthodologique.

Dans le chapitre suivant, nous allons considérer les différentes méthodes pour l'assistance intelligente à la mobilité des non-voyants ainsi que les algorithmes correspondants.

Chapitre 3 : Méthodes d'optimisation et stratégies algorithmiques pour les sous-systèmes d'assistance à la mobilité

3.1. Introduction

L'étude de la mobilité des personnes non-voyantes nécessite une approche progressive, allant de la compréhension des contraintes humaines et environnementales à la conception de solutions techniques adaptées. Dans cette continuité, les deux premiers chapitres ont posé les bases indispensables à l'élaboration d'un système d'assistance cohérent. Le premier a analysé les défis liés à la mobilité sans vision, ainsi que les limites des aides traditionnelles, mettant en évidence la complexité des situations de déplacement et la diversité des informations à traiter. Le deuxième chapitre a ensuite formalisé ces constats sous la forme de besoins fonctionnels et de modèles conceptuels, servant de cadre de référence pour la conception du système.

À partir de ces fondements, la problématique ne réside plus dans l'identification des besoins, mais dans leur traduction effective en mécanismes opérationnels capables de fonctionner en temps réel et dans des environnements dynamiques. Cette transition, qui constitue un passage essentiel de la théorie à la mise en œuvre, impose des choix méthodologiques rigoureux et des optimisations ciblées afin de concilier performance algorithmique, fiabilité du système et compatibilité avec les contraintes de l'usage réel.

Dans ce contexte, la conception d'un système d'assistance à la mobilité ne peut être envisagée comme une simple accumulation de modules indépendants. Elle requiert une organisation architecturale permettant l'interaction continue entre perception de l'environnement, interprétation de l'information et prise de décision. L'efficacité de l'assistance dépend ainsi de la capacité des différents sous-systèmes à coopérer de manière synchronisée, tout en préservant la stabilité et la cohérence globale du comportement du système.

Le présent chapitre s'inscrit dans cette logique de continuité méthodologique. Il est consacré à la description des méthodes, des optimisations et des choix de conception proposés pour répondre aux exigences identifiées précédemment. L'accent est mis sur la structuration du système en sous-systèmes fonctionnels, détection, évitement, localisation et classification, chacun étant conçu pour répondre à des objectifs spécifiques tout en contribuant à la performance globale de l'assistance.

Une attention particulière est portée à l'architecture générale du système, qui formalise les flux d'information et les mécanismes de coordination entre les différents modules. Cette architecture constitue le cadre dans lequel s'inscrivent les optimisations proposées, garantissant que les améliorations locales apportées à chaque sous-système s'intègrent harmonieusement dans une démarche globale orientée vers la mobilité autonome.

Les méthodes développées dans ce chapitre visent à améliorer la robustesse du système face à des environnements changeants, à renforcer la réactivité des mécanismes décisionnels et à assurer une interprétation plus structurée de l'environnement. Elles tiennent compte des contraintes

computationnelles inhérentes aux dispositifs embarqués, afin de garantir une mise en œuvre réaliste et compatible avec une utilisation prolongée en conditions réelles.

Ce chapitre constitue ainsi une étape charnière entre la modélisation théorique présentée précédemment et l'évaluation expérimentale du système. Il établit le cadre méthodologique et technique nécessaire à l'analyse des performances, à la validation des choix proposés et à la discussion critique des résultats.

3.2. Architecture générale du système

Le système d'assistance à la mobilité proposé repose sur une architecture modulaire et hiérarchisée, conçue pour assurer une coopération efficace entre les différentes fonctions nécessaires au déplacement autonome. Cette organisation vise à concilier robustesse algorithmique, réactivité en temps réel et évolutivité du système, tout en respectant les contraintes propres à une plateforme embarquée destinée à un usage quotidien. La figure 3.1 présente une vue d'ensemble de cette architecture globale, en mettant en évidence les principaux sous-systèmes fonctionnels ainsi que les flux d'information qui structurent leur interaction.

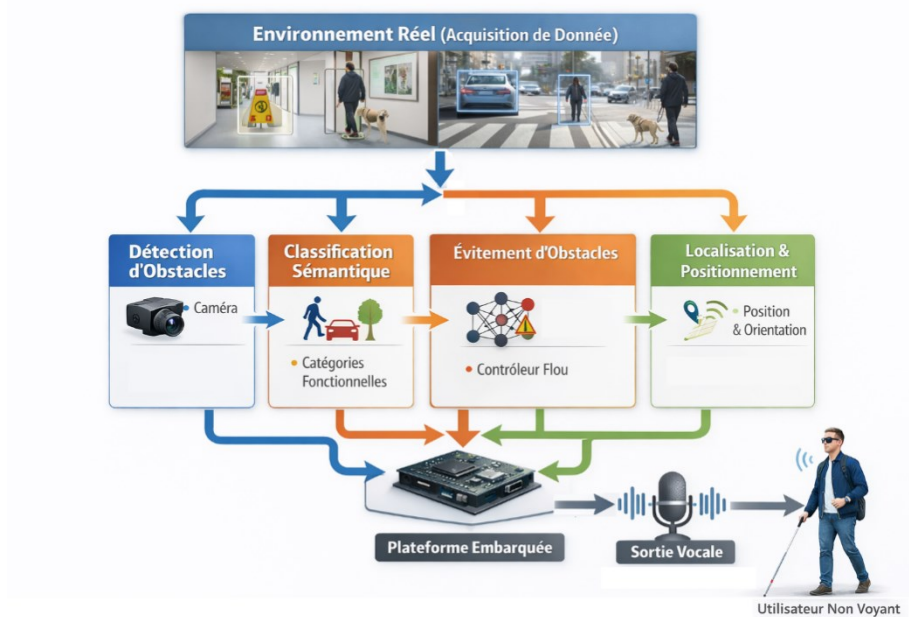


Figure 3.1. Architecture globale du système d'assistance à la mobilité pour personnes non-voyantes

L'architecture adoptée s'appuie sur une séparation claire des responsabilités fonctionnelles, permettant de structurer le traitement de l'information depuis l'acquisition des données jusqu'à leur restitution à l'utilisateur. Chaque sous-système remplit un rôle spécifique, tout en échangeant des informations avec les autres modules selon des flux définis. Cette approche favorise une intégration cohérente des

traitements, limite les dépendances directes entre composants et facilite l'optimisation locale des algorithmes sans remise en cause de l'ensemble du système.

- Le premier niveau de l'architecture est consacré à la perception de l'environnement. Le sous-système de détection des obstacles assure l'identification des éléments présents dans l'espace immédiat de déplacement à partir de données visuelles acquises en temps réel. Il repose sur des méthodes de vision par ordinateur et d'apprentissage profond, permettant d'extraire automatiquement des informations pertinentes à partir des images capturées. Ce module fournit une description initiale de la scène, indispensable à l'anticipation des situations potentiellement dangereuses et à la continuité du déplacement.
- En complément de cette détection, le sous-système de classification intervient pour enrichir l'interprétation des obstacles identifiés. Il attribue aux objets détectés des catégories fonctionnelles, en s'appuyant sur leurs caractéristiques visuelles et contextuelles. Cette étape permet de différencier les types de dangers et de fournir une information plus structurée aux modules décisionnels. La classification joue ainsi un rôle central dans l'adaptation des stratégies d'évitement et dans la pertinence des informations restituées à l'utilisateur.
- Le sous-système d'évitement constitue le noyau décisionnel de l'architecture. Il exploite conjointement les résultats de la détection, de la classification et les données issues de capteurs complémentaires afin de déterminer les ajustements nécessaires à la trajectoire. La logique d'évitement repose sur un contrôleur flou, choisi pour sa capacité à gérer des situations incertaines et évolutives sans imposer de comportements rigides. Cette approche permet de produire des décisions progressives et adaptées au contexte, garantissant une navigation fluide et sécurisée.
- Le sous-système de localisation complète l'architecture en fournissant une estimation continue de la position de l'utilisateur et en contribuant aux fonctions d'orientation à une échelle plus globale. Il permet d'inscrire les décisions locales d'évitement dans une compréhension spatiale plus large du déplacement, notamment lors de trajets vers des destinations prédéfinies. Ce module inclut également des mécanismes de partage de position, renforçant la sécurité de l'utilisateur dans des situations réelles d'usage.

L'ensemble de ces sous-systèmes est intégré sur une plateforme embarquée unique, assurant la synchronisation des traitements et la gestion des flux d'information. La restitution des résultats est réalisée principalement sous forme de messages vocaux, afin de fournir un retour clair et exploitable sans solliciter excessivement l'attention de l'utilisateur. Cette architecture globale constitue le cadre de

référence pour les méthodes d'optimisation et les stratégies algorithmiques détaillées dans les sections suivantes, chacune étant consacrée à l'analyse approfondie d'un sous-système spécifique.

3.3. Méthodologie du sous-système de détection

Le sous-système de détection des obstacles constitue le point d'entrée perceptif du système d'assistance à la mobilité. Il a pour fonction d'extraire, à partir des données visuelles acquises en temps réel, une description fiable et exploitable de l'environnement immédiat de l'utilisateur. Dans un contexte de navigation non visuelle, cette fonction joue un rôle déterminant, car elle conditionne la capacité du système à anticiper les situations dangereuses, à alimenter les mécanismes décisionnels et à fournir des informations pertinentes à l'utilisateur.

3.3.1. Détection d'obstacles en temps réel

La détection d'obstacles représente une fonction centrale du sous-système perceptif, car elle conditionne directement la sécurité, l'anticipation et la continuité du déplacement autonome des personnes non-voyantes. Contrairement à des applications classiques de vision artificielle orientées vers l'analyse hors ligne ou la reconnaissance statique, la navigation assistée impose des contraintes particulièrement strictes : le traitement doit être continu, à faible latence, robuste aux variations de l'environnement et compatible avec une exécution embarquée prolongée. Toute instabilité temporelle, retard de détection ou incohérence dans les résultats peut se traduire par un risque immédiat pour l'utilisateur.

Dans ce cadre, le choix de la méthode de détection ne peut être guidé uniquement par des critères de performance brute en laboratoire, mais doit résulter d'un compromis raisonné entre précision, réactivité, stabilité temporelle et coût computationnel. Les principales familles de méthodes de détection d'objets peuvent être classées en deux catégories : les approches à deux étapes (two-stage) et les approches à une seule étape (one-stage).

Les méthodes two-stage, telles que R-CNN [92], Fast R-CNN ou Faster R-CNN [93], reposent sur une séparation explicite entre la génération de régions candidates (Region Proposal Network) et la phase de classification et de régression des boîtes englobantes. Cette architecture hiérarchique permet d'atteindre une précision élevée, notamment dans des contextes contrôlés. Toutefois, elle introduit une complexité algorithmique importante, un temps de traitement variable selon la densité de la scène, et une latence globale incompatible avec les exigences de la mobilité en temps réel. Dans un scénario de déplacement, cette latence peut empêcher la détection suffisamment précoce d'un obstacle, rendant l'information inutilisable pour l'évitement.

À l'opposé, les approches à une seule étape, dont fait partie la famille YOLO (You Only Look Once) [94], proposent une reformulation complète du problème de détection. L'image est traitée globalement

par un réseau de neurones convolutionnel qui prédit, en une seule passe, l'ensemble des objets présents dans la scène comme illustré dans la figure 3.2. La détection est ainsi formulée comme un problème de régression directe, sans étape intermédiaire de proposition de régions.

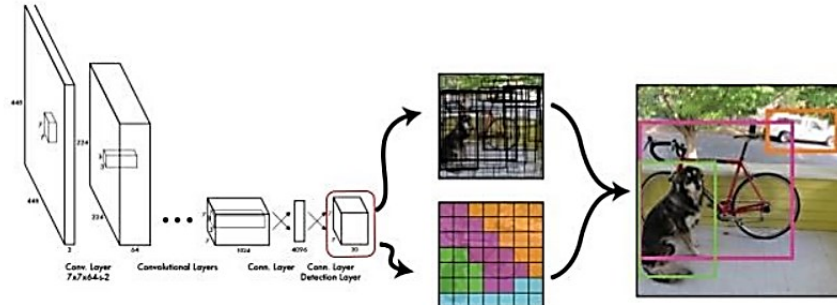


Figure 3.2. Principe de fonctionnement de YOLO

Mathématiquement, pour chaque cellule de la grille de prédiction, le réseau estime un ensemble de boîtes englobantes décrites par :

$$b = (x, y, w, h, c, p_1, \dots, p_K) \quad (1)$$

Où (x,y) représentent la position du centre de l'objet, (w,h) ses dimensions, « c » la probabilité de présence d'un objet, et « pk » les probabilités associées aux « K » classes. Cette formulation globale permet au modèle d'exploiter le contexte spatial complet de l'image, tout en garantissant un temps d'inférence constant et faible.

Ce principe constitue un avantage décisif pour l'assistance à la mobilité des personnes non-voyantes. Le déplacement s'effectue dans des environnements dynamiques, souvent encombrés, où les obstacles peuvent apparaître brutalement (piétons, vélos, trottinettes) ou être détectables seulement à courte distance (plots, bordures, sacs, petits objets au sol). Dans ces conditions, la priorité n'est pas uniquement la précision maximale, mais la capacité à fournir une information fiable et exploitable dans une fenêtre temporelle très réduite.

Les méthodes *one-stage*, et en particulier YOLO, présentent plusieurs atouts déterminants pour ce type d'application :

- Une latence faible et stable, essentielle pour la synchronisation avec la marche ;
- Une cohérence temporelle des détections, réduisant les fluctuations entre images successives ;
- Une architecture compacte, favorable à une intégration embarquée ;
- Une capacité à traiter des scènes complexes en temps réel, sans dépendance à des étapes intermédiaires coûteuses.

Ces éléments expliquent pourquoi l'approche YOLO est aujourd'hui largement privilégiée dans les systèmes embarqués critiques, tels que la robotique mobile, les véhicules autonomes ou les aides électroniques à la mobilité.

Le tableau 1 synthétise les différences fondamentales entre les approches *one-stage* et *two-stage* [95], dans une perspective orientée vers l'assistance à la mobilité.

Tableau 1. Comparaison entre méthodes de détection one-stage et two-stage

Critère	Méthodes one-stage	Méthodes two-stage
Principe	Régression directe globale	Propositions de régions et classification
Nombre d'étapes	Une seule	Deux étapes distinctes
Latence	Très faible et stable	Plus élevée et variable
Temps réel embarqué	Très adapté	Peu adapté
Complexité computationnelle	Modérée	Élevée
Robustesse en environnement dynamique	Élevée	Moyenne
Sensibilité à la densité de scène	Faible	Forte
Pertinence pour mobilité non visuelle	Très élevée	Limitée

Au regard de ces éléments, le choix de YOLO s'inscrit dans une logique méthodologique cohérente avec les objectifs du système d'assistance proposé. Il ne s'agit pas d'un choix opportuniste, mais d'une décision fondée sur les contraintes spécifiques de la mobilité sans vision, où la rapidité, la fiabilité et la continuité de l'information priment sur la recherche d'une précision maximale au détriment du temps de réponse. Les différentes versions proposées au fil des évolutions de l'algorithme traduisent des compromis distincts entre complexité architecturale, précision de détection et efficacité computationnelle. Dans le cadre d'un système d'assistance à la mobilité destiné pour un fonctionnement continu et embarqué, ces compromis doivent être évalués au regard des exigences de réactivité, de stabilité et de robustesse en environnement réel. La version YOLOv7 a ainsi été retenue comme base de développement, en raison de son équilibre entre performances de détection et coût de calcul, ainsi que de ses améliorations structurelles en matière d'agrégation multi-échelles et d'optimisation de l'inférence. Ce positionnement en fait un support pertinent pour l'élaboration d'un sous-système de détection capable de répondre aux contraintes temporelles de la navigation assistée, tout en offrant une marge d'optimisation ciblée pour la détection d'obstacles critiques. L'analyse de son fonctionnement et

de son architecture permet de mettre en évidence les apports de cette version, mais également ses limites dans le contexte spécifique de la mobilité non visuelle, qui motivent les adaptations proposées par la suite.

3.3.2. Fonctionnement et architecture de YOLOv7

YOLOv7 s'inscrit dans la lignée des algorithmes de détection d'objets one-stage, qui combinent en un seul passage la localisation et la classification des objets dans une image [96]. Ce principe repose sur le découpage de l'image en une grille, chaque cellule étant responsable de prédire la présence d'objets et leurs coordonnées. La détection est réalisée de manière end-to-end, garantissant des performances en temps réel tout en maintenant un compromis acceptable entre précision et vitesse de traitement. Cette architecture est particulièrement adaptée aux systèmes embarqués où le temps de latence est un critère critique, comme dans le cas des aides à la mobilité pour personnes non-voyantes.

L'architecture de YOLOv7 repose sur quatre blocs principaux : Input, Le réseau d'extraction de caractéristiques (backbone), Réseau d'agrégation multi-échelles (neck) et Le Module de prédiction (head).

- Input : Les images sont normalisées et redimensionnées pour correspondre aux contraintes du Le réseau d'extraction de caractéristiques.
- Le réseau d'extraction de caractéristiques : Il assure l'extraction des caractéristiques visuelles, en combinant des modules tels que le Convolutional Block Scheme (CBS), l'Extended Efficient Layer Aggregation Network (E-ELAN) et le module Maxpool 1 (MP1). Ces blocs permettent de capturer des informations contextuelles à différentes échelles tout en préservant la richesse des détails.
- Réseau d'agrégation multi-échelles : il intègre une structure de type Path Aggregation Feature Pyramid Network (PAFPN), combinée aux modules SPPCSPC et ELAN-H, afin de fusionner les caractéristiques extraites et les rendre exploitables pour la localisation et la classification.
- Le module de prédiction (head) : il est responsable de la prédiction finale, comprenant les coordonnées des boîtes englobantes, les scores de confiance et les classes. Il utilise notamment des blocs RepVGG (REP) et des convolutions 1×1 pour finaliser la détection.

La figure 3.3 illustre l'architecture complète de YOLOv7 et montre le flux de données depuis l'entrée jusqu'à la sortie des prédictions.

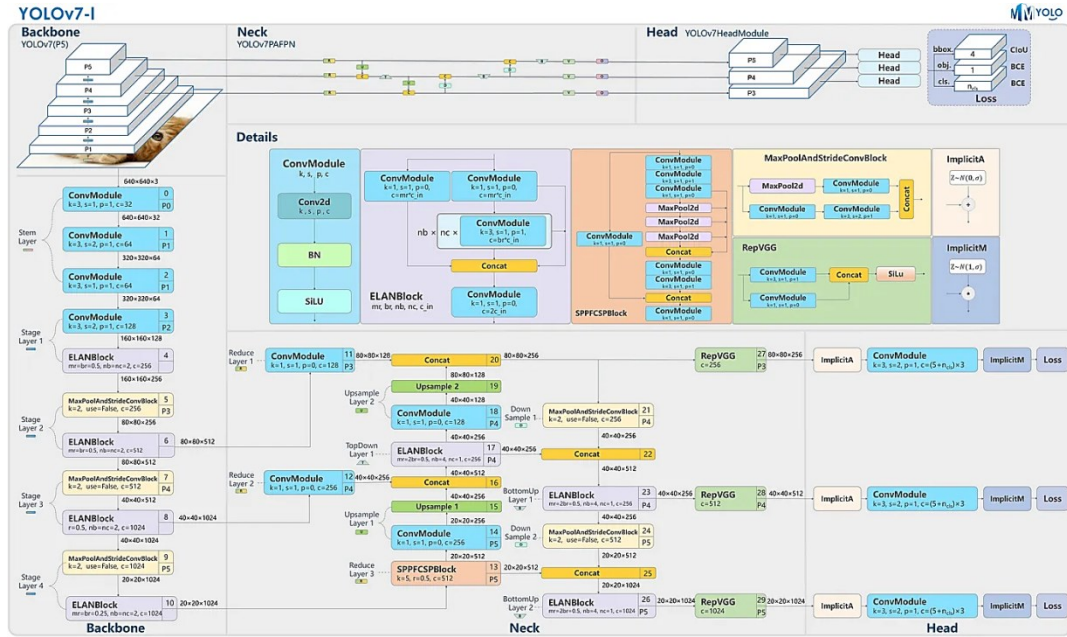


Figure 3.3. Architecture globale de YOLOv7

Malgré ses performances, YOLOv7 présente certaines limites qui peuvent représenter des risques pour les utilisateurs non-voyants. En particulier :

- Détection des objets proches et de grande taille : Les objets occupant une grande partie de l'image ou situés très près de l'utilisateur peuvent être mal localisés ou partiellement ignorés par le réseau, en raison de la structure de la grille et des étapes de downsampling dans le réseau d'extraction de caractéristiques.
- Détection des petits objets critiques : Des éléments tels que poteaux, bornes, boîtes de distribution, cônes de signalisation ou déchets sur le trottoir sont souvent mal détectés par YOLOv7 en condition réelle, en particulier lorsque leur contraste avec le fond est faible ou qu'ils sont partiellement occultés par d'autres objets.

Ces limitations constituent des enjeux majeurs pour la sécurité et la fluidité de la mobilité non-visuelle. Une détection manquée d'un petit obstacle à hauteur de pied ou un obstacle proche à hauteur de visage peut entraîner :

- Un risque de collision ou de chute,
- Une perte de trajectoire ou un ajustement brusque,
- Une perte de confiance dans le système et une surcharge cognitive pour l'utilisateur.

La prise en compte de ces problématiques a motivé l'optimisation de YOLOv7 dans notre méthodologie. L'objectif est d'améliorer la capacité du réseau à détecter de manière fiable les petits objets tout en

maintenant une réactivité en temps réel et une compatibilité avec les contraintes embarquées. Ainsi, cette version optimisée, que nous appelons YOLOv7-Plus, constitue la base du sous-système de détection et permet d'assurer que les informations critiques sont transmises rapidement et de manière fiable à l'utilisateur, améliorant directement la sécurité et la fluidité de la navigation.

3.3.3. Optimisation de YOLOv7 pour la détection d'obstacles

L'efficacité du sous-système de détection repose sur sa capacité à identifier, en temps réel et avec une fiabilité élevée, l'ensemble des obstacles présents dans l'environnement immédiat de l'utilisateur. Comme évoqué précédemment, la version standard de YOLOv7, bien que performante dans des contextes généraux, présente des limitations notables dans la détection des objets de petite taille et des obstacles proches ou de grande dimension, deux catégories critiques pour la sécurité des personnes non-voyantes. Un poteau fin, une borne, un petit objet sur le sol, ou à l'inverse un mur ou un véhicule surgissant à courte distance, représentent des dangers immédiats que le système doit impérativement anticiper.

La détection en temps réel pour l'assistance à la mobilité confronte l'algorithme à deux défis principaux aux conséquences symétriques :

- Les petits objets (poteaux, cônes, déchets) occupent peu de pixels dans l'image. Leur signal est faible et facilement perdu lors des réductions de résolution successives dans le Le réseau d'extraction de caractéristiques. Ils risquent de ne pas être activés dans les cartes de caractéristiques profondes, conduisant à de faux négatifs dangereux.
- Les objets proches ou de grande taille (murs, portes, véhicules à l'arrêt) peuvent occuper une part importante du champ de vision. Dans l'architecture standard, une grande boîte englobante peut être prédite de manière imprécise (mauvaise délimitation des bords) ou, dans certains cas, être fragmentée en plusieurs détections partielles en raison de la division de l'image en grilles. Cela génère des localisations inexacts ou des faux positifs multiples, brouillant l'interprétation de la scène.

Notre approche d'optimisation postule que ces deux problématiques trouvent leur origine dans des limitations liées à la représentation des caractéristiques à différentes échelles. Les modifications architecturales proposées visent donc à améliorer globalement la richesse, la préservation et la fusion de l'information visuelle à travers le réseau.

Les optimisations, conçues pour agir en synergie, portent sur deux composants clés : le réseau d'extraction de caractéristiques et le réseau d'agrégation multi-échelles.

3.3.3.1. Renforcement du réseau d'extraction de caractéristiques pour une extraction de caractéristiques enrichie et préservée

Le réseau d'extraction de caractéristiques (backbone) a été modifié pour préserver une information plus dense et plus fidèle dès les premières étapes de traitement, bénéficiant aussi bien aux petits objets qu'à la définition précise des grands.

- Augmentation de la profondeur des filtres dans les couches initiales : Accroître le nombre de filtres dans les premières couches convolutionnelles permet au réseau d'apprendre un dictionnaire plus complet et discriminant de caractéristiques de bas niveau (contours, textures, gradients). Pour les petits objets, cela renforce le signal initial avant qu'il ne s'affaiblisse. Pour les objets proches/grands, cela améliore la détection des contours fins et des transitions, cruciale pour une localisation aux bords nets.
- Ajustement des tailles de noyaux et des stratégies d'échantillonnage spatial : Un recalibrage des opérations de convolution et de downsampling permet d'optimiser le champ réceptif à différentes profondeurs. Cela aide à mieux contextualiser les petits objets dans leur environnement (en intégrant plus de contexte sans perdre le détail) et à maintenir une résolution spatiale suffisante pour définir avec précision l'étendue des objets larges.

3.3.3.2. Optimisation du réseau d'agrégation multi-échelles pour une fusion multi-échelles améliorée et équilibrée

Le réseau d'agrégation multi-échelles (PAFPN) est responsable de la fusion des caractéristiques à différentes résolutions. Son optimisation vise à créer des cartes de caractéristiques où l'information à toutes les échelles est représentée de manière équilibrée et discriminante.

- Raffinement des chemins de fusion et des connexions : En densifiant et en améliorant les connexions entre les couches de haute et basse résolution, on facilite la remontée des détails fins (essentiels pour les petits objets) vers les couches de prédiction, tout en maintenant l'influence des concepts sémantiques de haut niveau (utiles pour reconnaître et délimiter les grands objets).
- Affinage des modules de traitement dans le réseau d'agrégation multi-échelles : L'optimisation des blocs de convolution au sein du PAFPN permet de nettoyer et de renforcer les caractéristiques fusionnées. Cette étape réduit le bruit et améliore la distinction des entités, aidant à la fois à isoler un petit objet d'un fond complexe et à consolider la représentation cohérente d'un grand obstacle.

L'efficacité de YOLOv7-Plus proposé repose sur l'interaction étroite entre l'extracteur de caractéristiques et le réseau d'agrégation multi-échelles. L'extracteur de caractéristiques génère des cartes de caractéristiques initiales riches et discriminantes à partir de l'image d'entrée, tandis que le réseau d'agrégation multi-échelles les combine et les raffine afin de produire des représentations finales où :

- Les signaux des petits objets sont amplifiés et correctement contextualisés.
- Les caractéristiques des grands objets sont cohérentes spatialement et bien délimitées.

Le module de prédiction de détection, inchangé dans sa structure mais bénéficiant de ces entrées supérieures, peut alors générer des prédictions à la fois plus exhaustives (haut rappel) et plus précises géométriquement (haut IoU). Le Tableau 2 représente une comparaison entre l'architecture basique de YOLOv7 et celle optimisée YOLOv7-Plus proposé [97] et la figure 3.4 représente l'architecture de la version optimisée.

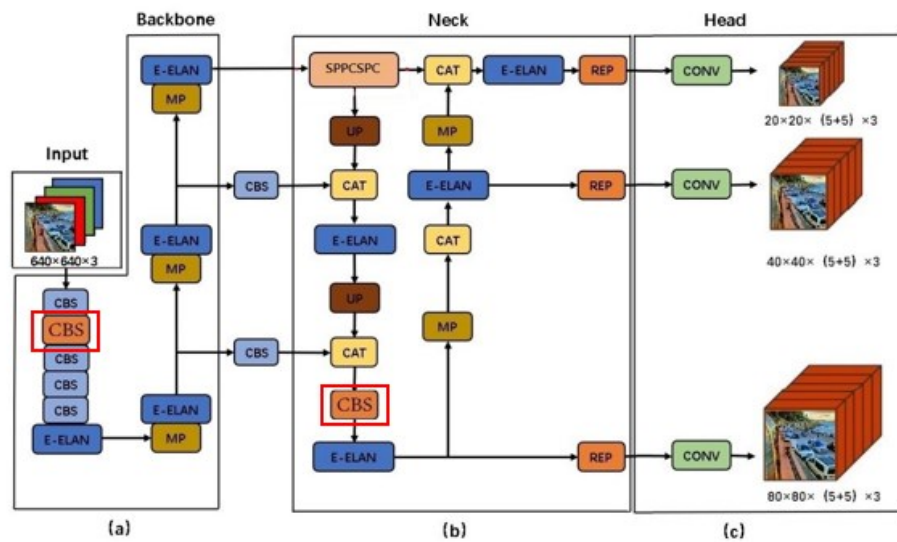


Figure 3.4. Architecture globale de la version optimisée

Tableau 2. Synthèse comparative des architectures YOLOv7 et YOLOv7-Plus proposé

Composant	YOLOv7	YOLOv7-Plus proposé	Impact sur la performance pour les petits objets	Impact sur la performance pour les objets proches/grands
Le réseau d'extraction de caractéristiques (backbone)	Structure CBS + E-ELAN standard	Profondeur de filtres augmentée (couches basses), noyaux et échantillonnage optimisés.	Préservation du signal faible, meilleur contexte.	Détection de contours plus précise, meilleure définition spatiale.
Réseau d'agrégation	Architecture standard	Chemins de fusion densifiés, modules	Meilleure remontée et distinction des détails fins.	Fusion cohérente des informations à plusieurs échelles pour

multi-échelles (neck)		de convolution affinés.		une représentation unifiée de l'objet.
Résultat global	Bonnes performances moyennes.	Extraction et fusion de caractéristiques améliorées.	↑ Rappel, ↓ Faux Négatifs.	↑ Précision de localisation (IoU), ↓ Faux Positifs multiples, détection plus stable.

3.3.3.3. Entraînement du modèle

Pour garantir la pertinence du modèle optimisé dans le contexte spécifique de l'assistance à la mobilité, son entraînement a été conduit sur le jeu de données Walk on the Road (WOTR) [69]. Ce corpus public, contenant 13 928 images annotées avec près de 190 000 objets, est spécifiquement conçu pour le développement d'aides à la navigation en milieu urbain. Il couvre 20 catégories d'objets réparties en obstacles typiques (arbre, poteau, cône, poubelle, borne, etc.) et indicateurs de cheminement (feux piétons, passages cloutés, pavés podotactiles, panneaux). Cette diversité, issue d'environnements réels variés (trottoirs, traversées, routes), assure une robustesse du modèle face aux scénarios complexes rencontrés par les utilisateurs.

Le protocole d'entraînement a intégré plusieurs pratiques visant à maximiser les performances tout en respectant les contraintes d'une plateforme embarquée :

- Pré-traitement et augmentation des données : Normalisation et techniques d'augmentation (recadrage, modification de luminosité/saturation) ont été appliquées pour améliorer la généralisation et la robustesse du modèle aux variations d'éclairage et d'angle.
- Apprentissage par transfert : Le modèle a été initialisé avec des poids pré-entraînés sur le large corpus MS COCO, accélérant la convergence et lui conférant une base de connaissances visuelles génériques solide.
- Réglage fin des hyperparamètres : Le taux d'apprentissage, le momentum et le weight decay ont été optimisés de manière adaptative pour favoriser un apprentissage stable et éviter le surajustement (overfitting), particulièrement crucial pour la détection des petits objets souvent sous-représentés.
- Optimisation mémoire : Des techniques telles que l'accumulation de gradients et un choix prudent de la taille de lot (batch size) ont permis de gérer efficacement la mémoire GPU durant l'entraînement, rendant le processus faisable sur des plateformes aux ressources limitées.

Le sous-système de détection, matérialisé par l'architecture optimisée YOLOv7-Plus, constitue la composante sensorielle fondamentale du système. Il fournit une cartographie fiable et exhaustive de l'environnement en temps réel, transformant le flux visuel brut en une représentation structurée des

éléments environnementaux. Cette capacité de perception précise, obtenue par l'optimisation ciblée du réseau d'extraction de caractéristiques et du réseau d'agrégation multi-échelles, représente la première étape indispensable de l'assistance à la mobilité.

Toutefois, la détection seule ne permet pas de guider activement l'utilisateur. Pour passer de la perception à l'action, l'information environnementale doit être interprétée et traduite en décisions de navigation. Cette fonction décisive de transformation de la description environnementale vers des instructions d'évitement est assumée par le sous-système d'évitement, dont l'architecture et les mécanismes sont détaillés dans la section suivante.

3.4. Méthodologie du sous-système de classification

Le sous-système de détection présenté précédemment permet d'identifier la présence d'obstacles. Toutefois, pour une assistance réellement efficace des personnes non-voyantes en environnement extérieur, il est indispensable d'enrichir cette information par une compréhension du comportement des obstacles, notamment de leur caractère statique ou dynamique. La distinction entre un obstacle immobile (bordure, poteau, véhicule stationné) et un obstacle en mouvement (piéton, cycliste, véhicule) constitue un élément clé de la prise de décision, car elle influe directement sur la stratégie d'évitement, le niveau d'alerte et le type de retour fourni à l'utilisateur. Un obstacle dynamique nécessite une anticipation de sa trajectoire future et une marge de sécurité plus importante, tandis qu'un obstacle statique peut être contourné selon une géométrie prévisible.

Dans ce contexte, cette section présente un sous-système de classification spatio-temporelle conçu pour analyser le comportement des obstacles détectés à partir de séquences vidéo courtes. L'approche proposée repose sur une architecture hybride combinant DINOv2 pour l'extraction de caractéristiques visuelles robustes et ConvLSTM pour la modélisation temporelle du mouvement, permettant ainsi une classification fiable des obstacles dans des environnements urbains complexes.

3.4.1. Principe de fonctionnement de ConvLSTM pour la modélisation temporelle

Le "Convolutional Long Short-Term Memory" (ConvLSTM) étend l'architecture LSTM classique aux données spatialement structurées en remplaçant les multiplications matricielles par des convolutions bidimensionnelles [98]. Cette adaptation est essentielle pour traiter des séquences d'images où la préservation de la structure spatiale est critique. Contrairement au LSTM standard qui opère sur des vecteurs aplatis, le ConvLSTM maintient la dimensionalité 2D des caractéristiques d'entrée, permettant ainsi de modéliser simultanément les dépendances spatiales et temporelles. Le fonctionnement interne d'une cellule ConvLSTM est régi par un système d'équations différentielles aux différences finies qui gèrent le flux d'information à travers trois portes principales :

La porte d'entrée, la porte d'oubli et la porte de sortie.

À chaque pas de temps « t », la cellule prend en entrée la carte de caractéristiques courante $X_t \in \mathbb{R}^{H \times W \times C}$ et produit un état caché $H_t \in \mathbb{R}^{H \times W \times C'}$ et un état de cellule $C_t \in \mathbb{R}^{H \times W \times C'}$.

Les équations fondamentales du ConvLSTM sont [99] :

$$i_t = \sigma(W_{x_i} * X_t + W_{h_i} * H_{t-1} + W_{c_i} \odot C_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{x_f} * X_t + W_{h_f} * H_{t-1} + W_{c_f} \odot C_{t-1} + b_f) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{x_c} * X_t + W_{h_c} * H_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_{x_o} * X_t + W_{h_o} * H_{t-1} + W_{c_o} \odot C_t + b_o) \quad (5)$$

$$H_t = o_t \odot \tanh(C_t) \quad (6)$$

Ou « * » représente l'opération de convolution 2D, $\odot$ désigne le produit élément par élément (Hadamard), « σ » est la fonction sigmoïde, « $\tanh$ » est la tangente hyperbolique, « W » sont des noyaux convolutifs apprenables, et « b » sont des termes de biais. La figure 3.5 illustre cette architecture.

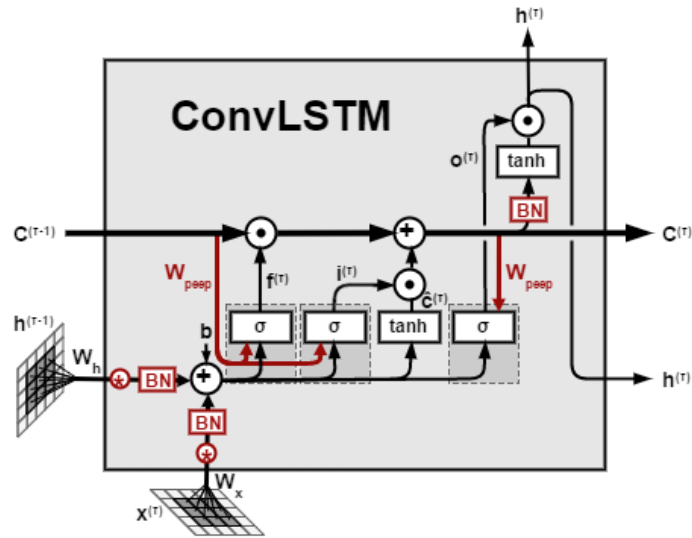


Figure 3.5. Architecture du ConvLSTM

Cette architecture permet de capturer des motifs spatiaux locaux via les convolutions tout en maintenant une mémoire temporelle à long terme grâce aux mécanismes de portes du LSTM, ce qui est particulièrement adapté à la modélisation des dynamiques comportementales dans les séquences vidéo.

3.4.2. Principe de fonctionnement de DINOv2 pour l'extraction de caractéristiques spatiales

L'analyse temporelle du mouvement repose fondamentalement sur la qualité des représentations visuelles extraites à chaque instant. Dans les environnements urbains réels, les scènes présentent des variations considérables d'éclairage (ensoleillement direct, ombres portées, conditions nocturnes), des occlusions partielles et des changements d'échelle (objets proches vs lointains). Les architectures convolutionnelles classiques, bien qu'efficaces, montrent leurs limites face à ces variations extrêmes.

DINOv2 (Distillation with NO labels) représente une avancée significative pour l'extraction de caractéristiques visuelles robustes [100]. Basé sur l'architecture Vision Transformer (ViT) [101], ce modèle est pré-entraîné de manière auto-supervisée sur un vaste corpus d'images suivant une méthodologie de distillation. La Figure 3.6 illustre cette architecture : un modèle pré-entraîné extrait des vecteurs de caractéristiques (feature vectors) qui sont ensuite traités par des couches linéaires avec une activation ReLU pour la classification finale.

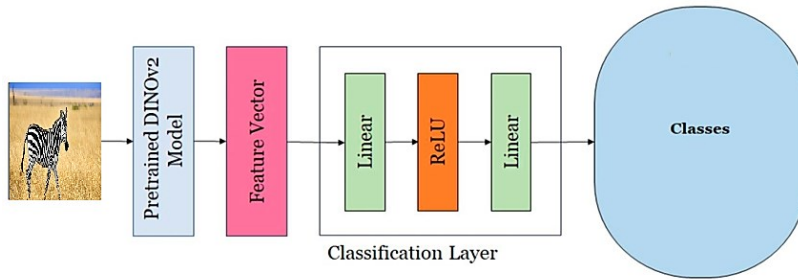


Figure 3.6. Architecture générale de DINOv2

Le principe fondamental de DINOv2 repose sur l'apprentissage par distillation auto-supervisée. Un modèle « enseignant » (*teacher*) guide l'apprentissage d'un modèle « étudiant » (*student*) sans recourir à des étiquettes manuelles. La fonction de perte est définie comme suit :

$$\mathcal{L} = - \sum_{x \in D} \sum_{k=1}^K P_t^k(x) \log P_s^k(x) \quad (7)$$

Où $P_t^k(x)$ et $P_s^k(x)$ sont les distributions de probabilité normalisées sur « K » classes de l'enseignant et de l'étudiant respectivement pour l'image « x » dans l'ensemble de données « D ».

L'architecture Vision Transformer procède en divisant l'image d'entrée « $I \in \mathbb{R}^{H \times W \times 3}$ » en une séquence de « N » patches « $P \times P$ ». Chaque patch est transformé linéairement en un vecteur de dimension « D » :

$$z_i^0 = \text{Linear}(\text{Flatten}(X_i)) + p_i \quad (8)$$

Où « p_i » est un vecteur de position apprenable qui encode l'information spatiale. Ces tokens sont ensuite traités par « L » blocs Transformer, chacun contenant un mécanisme d'attention multi-têtes [102] :

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (9)$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (10)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (11)$$

L'entraînement auto-supervisé confère à DINOv2 plusieurs propriétés avantageuses pour les applications de navigation en environnement réel : une invariance accrue aux transformations, une forte généralisation hors domaine et l'extraction de représentations sémantiques riches. Ces caractéristiques en font un encodeur visuel particulièrement adapté aux conditions variables rencontrées en milieu urbain. Dans l'architecture présentée (Figure 3.6), les caractéristiques extraites par DINOv2 sont ensuite traitées par des couches linéaires avec activation ReLU pour une adaptation spécifique à la tâche. Cette combinaison permet d'exploiter les représentations riches et robustes de DINOv2 tout en les orientant vers les spécificités de la classification comportementale.

3.4.3. Méthodologie hybride proposée

L'analyse de séquences vidéo pour la classification comportementale des obstacles impose une modélisation conjointe des informations spatiales et temporelles. Dans un contexte d'assistance à la mobilité des personnes non-voyantes, cette exigence est d'autant plus critique que le système doit non seulement détecter les obstacles, mais également anticiper leur comportement afin de garantir des décisions sûres et adaptées.

Les modèles de type vision transformer auto-supervisés, tels que DINOv2, sont particulièrement performants pour l'extraction de représentations spatiales discriminantes et invariantes face aux variations d'illumination, de perspective et de bruit visuel. Toutefois, ces modèles opèrent de manière indépendante sur chaque trame et ne capturent pas explicitement les dépendances temporelles entre images successives. À l'inverse, les réseaux ConvLSTM sont spécifiquement conçus pour modéliser les dynamiques spatio-temporelles, mais leur efficacité repose fortement sur la qualité et la stabilité des caractéristiques d'entrée.

La méthodologie proposée repose sur la fusion de ces deux approches complémentaires comme montré dans la figure 3.7. DINOv2 est utilisé comme extracteur spatial robuste, fournissant des représentations visuelles normalisées, tandis que le ConvLSTM exploite ces représentations pour apprendre les évolutions temporelles et comportementales des obstacles. Cette association permet de dissocier l'apprentissage des invariances spatiales de la modélisation dynamique, améliorant ainsi la fiabilité globale du système.

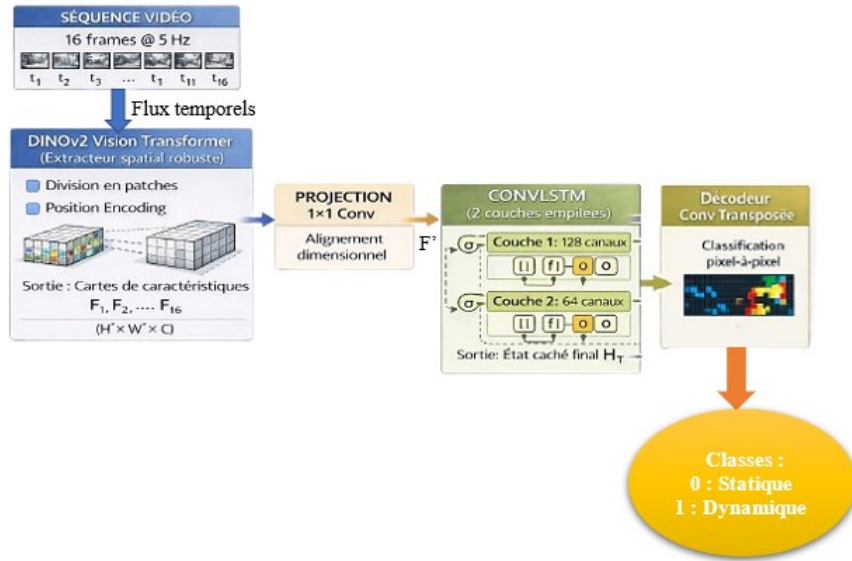


Figure 3.7. Architecture de la méthode hybride proposée pour la classification

L'architecture hybride adoptée suit un pipeline séquentiel modulaire, structuré en quatre blocs fonctionnels principaux :

- **Extraction spatiale par DINOv2** : Chaque trame vidéo « I_t » est traitée indépendamment par le réseau DINOv2 afin d'extraire une carte de caractéristiques spatiales $F_t \in \mathbb{R}^{H' \times W' \times C}$. Ces cartes encodent des informations sémantiques de haut niveau, issues d'un pré-entraînement auto-supervisé à grande échelle. Cette étape assure une forte robustesse aux variations photométriques et géométriques, tout en fournissant une représentation cohérente et stable pour l'analyse temporelle ultérieure.
- **Projection adaptative des caractéristiques** : Afin d'adapter les sorties de DINOv2 au module temporel, une couche convolutive 1×1 est appliquée à chaque carte F_t . Cette projection permet d'aligner la dimensionnalité des canaux, de réduire la redondance informationnelle et de normaliser les distributions des caractéristiques. Elle joue ainsi un rôle d'interface entre l'extracteur spatial et le modélisateur temporel.
- **Modélisation spatio-temporelle par ConvLSTM** : Les cartes projetées sont ensuite organisées en une séquence temporelle : $\{F_1, F_2, \dots, F_N\}$, où « N » représente le nombre de trames constituant une fenêtre temporelle vidéo. Cette séquence est traitée par un réseau ConvLSTM composé de deux couches empilées (128 puis 64 canaux). La première couche est dédiée à la capture des dynamiques locales et transitoires, telles que les mouvements rapides ou les changements brusques de trajectoire. La seconde couche intègre ces informations sur un horizon temporel plus long, permettant de modéliser des comportements plus globaux et persistants. Dans l'implémentation retenue, la fenêtre temporelle est fixée à « $N=16$ » trames, extraites à une

fréquence d'échantillonnage de 5 images par seconde, correspondant à une durée d'observation de 3,2 secondes. Cette structuration hiérarchique favorise une compréhension progressive des interactions spatio-temporelles, essentielle pour la classification comportementale des obstacles.

- **Décodeur et sortie comportementale :** Les états cachés du ConvLSTM sont décodés à l'aide de convolutions transposées, conduisant à la génération d'une carte dense de segmentation comportementale : $M \in \{0,1\}^{256 \times 256}$. Une fonction d'activation sigmoïde est appliquée en sortie afin de produire une interprétation probabiliste binaire, directement exploitable par les modules décisionnels d'assistance à la navigation.

L'entraînement du modèle hybride est réalisé à partir de séquences issues de la base de données BDD100K, qui regroupe un large ensemble de vidéos acquises en conditions réelles et couvrant une grande diversité de scènes, de conditions météorologiques et d'environnements dynamiques. Le processus d'apprentissage est structuré en deux phases. Dans une première phase, les poids de DINOv2 sont maintenus figés, tandis que le ConvLSTM et les couches de projection sont entraînés afin de stabiliser l'apprentissage temporel. Dans une seconde phase, un affinement global est effectué avec un taux d'apprentissage réduit, permettant une adaptation conjointe des représentations spatiales et temporelles sans compromettre les propriétés d'invariance acquises. Des techniques d'augmentation des données, incluant des variations photométriques, du flou cinétique et des occlusions synthétiques, sont appliquées afin d'améliorer la robustesse du modèle. La fonction de perte repose sur une combinaison pondérée de pertes binaires, avec une pénalisation renforcée des faux négatifs, conformément aux exigences de sécurité des systèmes d'assistance aux personnes non-voyantes.

La fusion DINOv2–ConvLSTM confère au système plusieurs avantages techniques majeurs. Elle améliore la robustesse perceptuelle face aux artefacts d'acquisition, tout en permettant une modélisation temporelle fine compatible avec des contraintes de temps réel. La séparation explicite entre extraction spatiale et modélisation dynamique favorise également une meilleure généralisation vers des environnements non observés lors de l'apprentissage.

Enfin, la sortie structurée sous forme de carte pixel-à-pixel constitue une interface directe avec les modules d'évitement et de localisation, assurant une intégration cohérente au sein de la chaîne globale d'assistance à la navigation, où la précision spatiale est enrichie par une conscience contextuelle temporelle.

3.5. Méthodologie du sous-système d'évitement d'obstacles

Après la phase de détection, l'évitement des obstacles constitue la fonction décisionnelle opérationnelle du système d'assistance à la mobilité. Il assure la transformation des informations environnementales, souvent incertaines et évolutives, en actions correctrices permettant d'adapter le déplacement de l'utilisateur en temps réel. Dans un contexte de navigation non visuelle, cette fonction revêt une importance critique, car toute décision inadéquate ou tardive peut compromettre la sécurité, la continuité du déplacement et la confiance de l'utilisateur dans le dispositif.

Le sous-système d'évitement est ainsi conçu comme un mécanisme réactif, agissant à l'échelle locale, et destiné à gérer les situations de proximité immédiate avec les obstacles. Il s'inscrit dans une chaîne fonctionnelle continue reliant la perception à la localisation, et contribue à la stabilité globale du comportement du système.

3.5.1. Évitement des obstacles en temps réel

Dans le domaine de l'assistance à la mobilité des personnes non-voyantes, l'évitement des obstacles repose principalement sur des approches décisionnelles locales, capables de produire des réponses immédiates à partir des informations issues des capteurs. Parmi ces approches, les contrôleurs basés sur la logique floue constituent la solution la plus largement adoptée en pratique, en raison de leur capacité à gérer l'imprécision, la variabilité des données sensorielles et les transitions rapides de l'environnement.

L'évitement en temps réel repose sur la nécessité d'interpréter des informations partielles, souvent bruitées, relatives à la distance, à la configuration spatiale et à la dynamique des obstacles. Les méthodes classiques fondées sur des seuils déterministes présentent des limites importantes dans ce contexte, car elles conduisent à des comportements discontinus et peu adaptés à la complexité des scènes réelles. À l'inverse, la logique floue permet de modéliser ces informations sous forme de degrés continus d'appartenance, offrant ainsi une représentation plus fidèle des situations rencontrées lors d'un déplacement réel.

Dans cette approche, les décisions d'évitement ne résultent pas d'un choix binaire, mais d'une inférence graduelle, prenant en compte plusieurs niveaux de proximité et de risque. Cette continuité décisionnelle est essentielle pour produire des ajustements progressifs de la trajectoire, compatibles avec la cinématique de la marche humaine et les capacités de réaction de l'utilisateur. Elle contribue également à réduire les comportements brusques ou instables, susceptibles d'augmenter la charge cognitive ou de générer une perte de confiance dans le système.

Le choix de la logique floue comme mécanisme central d'évitement dans ce travail est motivé par plusieurs considérations méthodologiques. D'une part, elle offre une robustesse intrinsèque face aux

incertitudes de mesure, particulièrement adaptée aux capteurs de distance utilisés en mobilité embarquée. D'autre part, elle permet une formalisation explicite du raisonnement décisionnel, à travers un ensemble de règles interprétables, facilitant l'analyse, l'adaptation et l'extension du système. Enfin, le coût computationnel modéré des contrôleurs flous garantit une exécution en temps réel, condition indispensable pour une assistance continue et fiable.

Ainsi, l'évitement proposé dans ce travail s'appuie sur une stratégie floue réactive, conçue pour fournir des décisions stables, progressives et adaptées aux contraintes de la mobilité non visuelle.

3.5.2. Principe général du contrôleur flou

Un contrôleur flou est un système de décision non linéaire fondé sur la théorie des ensembles flous, permettant de formaliser un raisonnement approchant à partir d'informations imprécises ou incertaines. Contrairement aux approches déterministes classiques, il ne nécessite pas de modèle mathématique exact du système à contrôler et intègre directement l'incertitude des données d'entrée dans le processus décisionnel.

Le fonctionnement général d'un contrôleur flou repose sur quatre étapes fondamentales : la fuzzification, l'inférence floue, l'agrégation des règles et la défuzzification [103]. Cette architecture assure une séparation explicite entre la représentation des connaissances, le mécanisme de raisonnement et la génération de la commande.

La fuzzification consiste à projeter les variables d'entrée numériques $x \in \mathbb{R}$, issues des capteurs, vers des ensembles flous « A_i » à l'aide de fonctions d'appartenance $\mu_{A_i}(x)$, définies par :

$$\mu_{A_i} : \mathbb{R} \rightarrow [0, 1] \quad (12)$$

Ces fonctions permettent de représenter des notions qualitatives telles que faible, moyenne ou élevée, tout en conservant une transition continue entre les états. Les formes les plus couramment utilisées sont triangulaires, trapézoïdales ou gaussiennes, en raison de leur simplicité et de leur robustesse face au bruit. L'inférence floue repose sur un ensemble de règles linguistiques de type :

$$R_k: \text{Si } x_1 \text{ est } A_1^k \text{ et } x_2 \text{ est } A_2^k \text{ alors } y \text{ est } B^k$$

Le degré d'activation de chaque règle est calculé à l'aide d'opérateurs logiques flous, généralement basés sur le minimum ou le produit :

$$\alpha_k = \min(\mu_{A_1^k}(x_1), \mu_{A_2^k}(x_2, \dots)) \quad (13)$$

$$\text{Où :} \quad \alpha_k = \prod_i \mu_{A_i^k}(x_i) \quad (14)$$

Ces degrés d'activation pondèrent les conclusions floues associées à chaque règle, traduisant ainsi l'influence relative de chaque situation sur la décision finale.

Les sorties partielles sont ensuite combinées lors de l'agrégation, afin d'obtenir un ensemble flou global $\mu_B(y)$, généralement défini par :

$$\mu_B(y) = \max_k(\alpha_k \cdot \mu_{B^k}(y)) \quad (15)$$

Cette étape garantit une prise de décision cohérente en intégrant l'ensemble des règles activées.

La défuzzification permet enfin de transformer l'ensemble flou de sortie en une valeur numérique exploitable par le système. La méthode la plus utilisée est celle du centre de gravité, définie par :

$$y^* = \frac{\int y \cdot \mu_B(y) dy}{\int \mu_B(y) dy} \quad (16)$$

Cette formulation assure une sortie continue et stable, évitant les changements brusques de commande susceptibles de perturber le déplacement de l'utilisateur.

Dans les systèmes d'assistance à la mobilité des non-voyants, le contrôleur flou se distingue par sa capacité à produire des décisions progressives, interprétables et compatibles avec les contraintes du temps réel. Ces caractéristiques expliquent son recours fréquent à l'évitement des obstacles dans un environnement dynamique.

3.5.3. Contrôleur flou proposé pour l'évitement en temps réel

Dans un système d'assistance à la mobilité destiné aux personnes non-voyantes, la simple connaissance de la distance aux obstacles ne suffit pas à garantir une navigation sûre et fluide. En environnement réel, les situations à risque émergent non seulement de la proximité spatiale, mais aussi de la nature de l'obstacle, de son comportement dynamique et de son interaction potentielle avec l'utilisateur. Dans ce contexte, le processus d'évitement doit dépasser une logique purement géométrique pour intégrer une interprétation sémantique de la scène, issue des sous-systèmes de détection et de classification.

La contribution proposée dans cette thèse s'inscrit précisément dans cette perspective. Contrairement à l'approche préliminaire présentée dans nos travaux antérieurs, reposant sur deux capteurs de distance uniquement [104], l'architecture développée ici étend le schéma décisionnel à trois capteurs de distance de type ToF VL53L1X, tout en fusionnant ces mesures avec les informations sémantiques fournies par les sous-systèmes de détection et de classification comme illustré dans la figure 3.8.

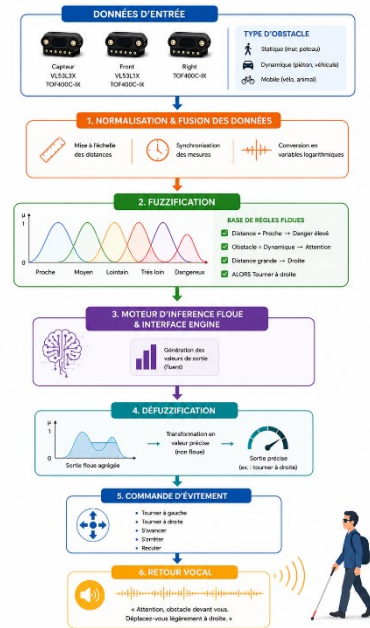


Figure 3.8. Organigramme du contrôleur flou pour évitement d'obstacle

Cette fusion permet au contrôleur flou de raisonner non seulement sur la proximité des obstacles, mais également sur leur catégorie (statique ou dynamique), en accordant une priorité systématique aux obstacles dynamiques, considérés comme les plus dangereux pour l'utilisateur :

3.5.3.1. Chaîne décisionnelle globale et cohérence inter-sous-systèmes

Le fonctionnement du contrôleur flou proposé repose sur une chaîne décisionnelle cohérente, structurée en quatre étapes successives :

- **Détection visuelle de l'obstacle :** Le sous-système de détection identifie la présence d'un obstacle dans le champ de vision de la caméra embarquée et fournit sa localisation relative dans l'image.
- **Classification sémantique :** Une fois l'obstacle détecté, le sous-système de classification détermine sa nature : Obstacle statique (poteau, mur, borne, mobilier urbain, etc.) et obstacle dynamique (piéton, vélo, trottinette, animal, etc.).
- **Estimation de la distance par capteurs ToF :** La localisation spatiale de l'obstacle est ensuite affinée à l'aide de trois capteurs de distance ToF VL53L1X, positionnés respectivement à gauche, au centre et à droite du dispositif. Cette configuration permet : une couverture latérale étendue, une estimation plus robuste de la distance réelle, une réduction de l'ambiguïté liée aux mesures isolées.
- **Prise de décision par le contrôleur flou :** Le contrôleur flou exploite simultanément les distances mesurées par les trois capteurs et la nature de l'obstacle (statique ou dynamique), afin

de générer une décision d'évitement adaptée, en privilégiant systématiquement l'évitement de l'obstacle dynamique le plus proche, lorsque plusieurs obstacles sont présents.

Cette organisation garantit une continuité logique entre la perception, l'interprétation et l'action, tout en maintenant une latence compatible avec un déplacement naturel.

3.5.3.2. Variables d'entrée et de sortie du contrôleur flou

Le contrôleur flou proposé est défini par quatre variables d'entrée et une variable de sortie principale.

- **Variables d'entrée :** Capteur gauche (CG), Capteur frontal (CF), Capteur droit (CD) et Type d'obstacle (TO).

Les trois variables fournies par les capteurs ToF sont exprimées en termes linguistiques : Proche, Moyenne, Lointaine. La variable linguistique issue du sous-système de classification est définie par deux modalités : Statique et Dynamique. La figure 3.9 présente les variables linguistiques de l'entrée du sous-système d'évitement.

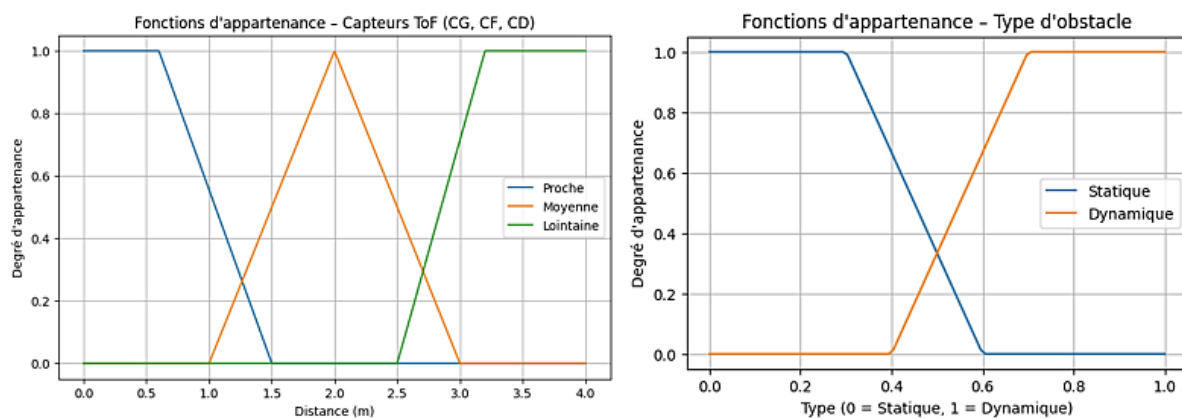


Figure 3.9. Fonctions d'appartenance des variables linguistiques d'entrée

- **Variable de sortie :** Décision d'évitement (DE) définie par les modalités suivantes : Évitement à gauche, Évitement à droite, Ralentissement / arrêt, Maintien de trajectoire et Faire demi-tour.

3.5.3.3. Base de règles floues

La base de règles constitue le cœur de la contribution proposée. Elle est conçue pour refléter un principe fondamental : « en présence d'un obstacle dynamique, la décision d'évitement est prioritaire, indépendamment de la configuration des obstacles statiques ». Le tableau 3 illustre un extrait représentatif de la base de règles floues intégrant les quatre entrées du système.

Tableau 3. Base de règles floues étendue avec gestion des situations critiques

CG	CF	CD	Type d'obstacle	Décision d'évitement
Proche	Proche	Loin	Dynamique	Évitement à droite
Loin	Proche	Proche	Dynamique	Évitement à gauche
Moyen	Proche	Moyen	Dynamique	Ralentissement / arrêt
Proche	Moyen	Loin	Statique	Évitement à droite
Loin	Moyen	Proche	Statique	Évitement à gauche
Moyen	Loin	Moyen	Statique	Maintien de trajectoire
Proche	Proche	Proche	Dynamique	Faire demi-tour
Proche	Proche	Proche	Statique	Faire demi-tour
Loin	Loin	Loin	Dynamique	Maintien de trajectoire
Loin	Loin	Loin	Statique	Maintien de trajectoire

L'introduction de la décision « Faire demi-tour » répond à des situations de blocage spatial total, caractérisées par la détection simultanée d'un obstacle proche sur les trois axes (gauche, frontal et droit). Dans un tel contexte, aucune manœuvre latérale ne garantit un niveau de sécurité suffisant, rendant le demi-tour la stratégie la plus fiable pour préserver l'intégrité de l'utilisateur. Cette règle est appliquée aussi bien aux obstacles dynamiques qu'aux obstacles statiques, bien que le risque associé aux obstacles dynamiques demeure plus élevé en raison de leur caractère imprévisible. À l'inverse, lorsque les trois capteurs indiquent une distance lointaine, l'environnement est considéré comme dégagé. Dans ce cas, que l'obstacle soit statique ou dynamique, la décision « Maintien de trajectoire » est privilégiée, traduisant une situation de navigation sûre ne nécessitant aucune adaptation du déplacement. Cette règle contribue à éviter des réactions excessives du système et garantit une navigation fluide et naturelle.

Au-delà des réactions locales aux obstacles, l'assistance à la mobilité requiert une gestion cohérente de la position et de l'orientation de l'utilisateur. Le sous-système de localisation, présenté dans la section suivante, répond à cette exigence en assurant le suivi spatial du déplacement.

3.6. Méthodologie du sous-système de localisation

Le sous-système de localisation est conçu pour fournir une estimation précise et continue de la position et de l'orientation de l'utilisateur en environnement extérieur. Cette fonctionnalité est essentielle pour contextualiser spatialement les informations de détection d'obstacles et pour permettre une navigation assistée vers des destinations prédéfinies. La spécificité de l'assistance à la mobilité impose des

contraintes strictes en matière de précision, de latence et de robustesse, particulièrement dans les environnements urbains où les signaux de positionnement peuvent être dégradés.

Pour répondre à ces exigences, notre approche repose sur une architecture de fusion de capteurs basée sur un filtre de Kalman étendu (EKF). Cette méthode permet d'intégrer de manière optimale les données d'une unité de mesure inertielle (IMU) et d'un récepteur GNSS, en exploitant leurs complémentarités pour compenser leurs limitations respectives.

3.6.1. Principes théoriques du filtre de Kalman étendu pour l'estimation d'état

La fusion des données issues de l'unité de mesure inertielle (IMU) et du récepteur GNSS soulève un problème d'estimation d'état complexe, principalement en raison des différences de fréquence d'échantillonnage, des incertitudes de mesure et des non-linéarités associées à la dynamique du mouvement. À ces contraintes s'ajoutent des exigences strictes de fonctionnement en temps réel et de maîtrise de la charge de calcul, imposées par une implémentation sur plateforme embarquée. Dans ce contexte, le filtre de Kalman étendu (Extended Kalman Filter – EKF) constitue un cadre théorique et algorithmique particulièrement adapté à la fusion de capteurs pour les applications de navigation assistée.

Le filtre de Kalman est un algorithme d'estimation récursive probabiliste visant à estimer l'état interne d'un système dynamique à partir d'un modèle d'évolution et de mesures bruitées [105]. Pour les systèmes linéaires soumis à des bruits gaussiens, il fournit une estimation optimale au sens de la minimisation de l'erreur quadratique moyenne. Son principe repose sur l'alternance de deux phases complémentaires : une phase de prédiction, durant laquelle l'état est propagé à l'aide du modèle dynamique, et une phase de correction, dans laquelle les observations disponibles sont intégrées afin d'affiner cette prédiction. L'estimation résultante découle d'une pondération adaptative entre le modèle et les mesures, pondération directement gouvernée par leurs matrices de covariance respectives.

Dans les applications réelles de navigation, les équations décrivant l'évolution de l'état et les relations reliant les mesures à cet état sont généralement non linéaires. Le filtre de Kalman étendu permet de traiter ce cas en procédant à une linéarisation locale des modèles autour de l'estimation actuelle [106]. Cette approximation est réalisée par le calcul des matrices jacobiennes des fonctions de transition et d'observation, ce qui rend possible l'application du formalisme du filtre de Kalman tout en conservant son caractère récursif. Grâce à cette approche, l'EKF offre un compromis équilibré entre précision de l'estimation, robustesse face aux incertitudes de mesure et complexité de calcul maîtrisée. Ces propriétés en font un outil de référence pour la localisation embarquée, en particulier dans le cadre de l'assistance à la mobilité des personnes non-voyantes, où la continuité et la fiabilité de l'estimation de position conditionnent directement la sécurité du déplacement. Dans le cadre de la localisation, le système dynamique est représenté par deux équations fondamentales.

L'équation d'état décrit l'évolution temporelle du système :

$$x_k = f(x_{k-1}, u_k) + w_k \quad (17)$$

L'équation de mesure relie l'état du système aux observations disponibles :

$$z_k = h(x_k) + v_k \quad (18)$$

Dans ces équations, « x_k » désigne le vecteur d'état à l'instant « k », « u_k » représente le vecteur d'entrée de commande, et « z_k » correspond au vecteur de mesure. La fonction $f(\cdot)$ modélise la dynamique du système, tandis que « $h(\cdot)$ » décrit la relation entre l'état et les mesures. Les termes « w_k » et « v_k » représentent respectivement le bruit de processus, de covariance « Q_k », et le bruit de mesure, de covariance « R_k », tous deux supposés blancs gaussiens.

▪ Phase de prédiction

Cette étape projette l'état estimé et son incertitude associée depuis l'instant « $k-1$ » vers l'instant k , en utilisant exclusivement le modèle dynamique. L'état prédit « $\hat{x}_{k|k-1}$ » est obtenu par application de la fonction de transition :

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_k) \quad (19)$$

La covariance de l'erreur de prédiction « $P_{k|k-1}$ » est calculée par propagation de l'incertitude :

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \quad (20)$$

Où $F_k = \left. \frac{\partial f}{\partial x} \right|_{\hat{x}_{k-1|k-1}}$ est la matrice jacobienne de la fonction de transition, évaluée à l'estimation courante, $P_{k-1|k-1}$ et représente la covariance de l'erreur d'estimation à l'instant précédent.

▪ Phase de correction

Lorsqu'une mesure « z_k » devient disponible, elle est intégrée pour corriger l'état prédit. L'innovation « $\tilde{y}_k$ », différence entre la mesure observée et la mesure prédite, est calculée par :

$$\tilde{y}_k = z_k - h(\hat{x}_{k|k-1}) \quad (21)$$

La covariance de l'innovation « S_k » est déterminée par :

$$S_k = H_k P_{k|k-1} H_k^T + R_k \quad (22)$$

Où $H_k = \left. \frac{\partial h}{\partial x} \right|_{\hat{x}_{k|k-1}}$ est la matrice jacobienne de la fonction d'observation. Le gain de Kalman « K_k », qui pondère l'influence de l'innovation sur la correction, est donné par :

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (23)$$

L'état estimé final $\hat{x}_{k|k}$ et sa covariance associée $P_{k|k}$ sont alors obtenus par :

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \quad (24)$$

$$P_{k|k} = (I - K_k H) P_{k|k-1} \quad (25)$$

Ce cycle prédiction-correction, exécuté récursivement à chaque nouvelle mesure, permet une estimation optimale et continue de l'état du système.

3.6.2. Approche de localisation proposée basée sur la fusion IMU-GPS

La localisation continue de l'utilisateur en environnement extérieur repose sur la fusion cohérente de capteurs aux propriétés complémentaires. Dans cette optique, l'approche proposée s'appuie sur l'intégration conjointe d'une unité de mesure inertielle (IMU) et d'un récepteur GNSS, au sein d'un filtre de Kalman étendu (EKF). La figure 3.10 illustre l'architecture proposé.

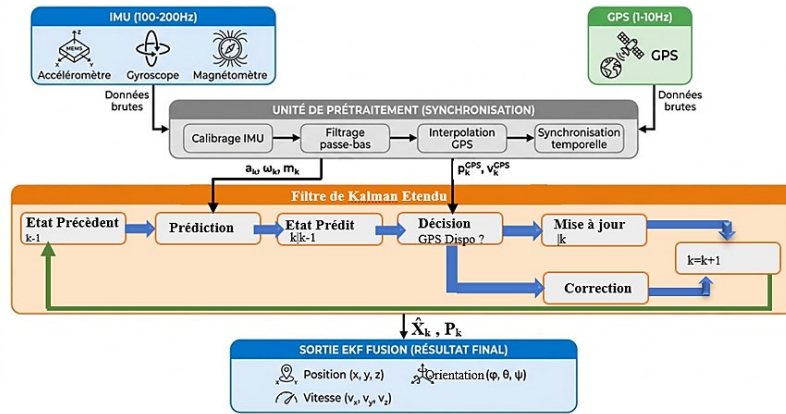


Figure 3.10. Architecture de fusion IMU-GPS par Filtre de Kalman Etendu

L'IMU, composée d'un accéléromètre et d'un gyroscope triaxiaux, fournit des mesures de mouvement à haute fréquence (typiquement entre 100 et 200 Hz). Ces capteurs permettent de suivre finement la dynamique instantanée de l'utilisateur (accélérations, rotations, variations d'orientation), ce qui est essentiel pour assurer la continuité de la localisation, notamment lors des phases de déplacement rapide ou lorsque les signaux GNSS sont indisponibles. Cependant, les mesures inertielles étant intégrées dans le temps, elles sont sujettes à une dérive cumulative, principalement due aux biais et au bruit des capteurs. À l'inverse, le récepteur GNSS fournit une estimation absolue de la position à une fréquence plus faible (1 à 10 Hz), mais avec une référence géographique globale. Ces mesures sont peu affectées par la dérive temporelle, mais peuvent être fortement dégradées en milieu urbain dense, sous couvert végétal ou à proximité de bâtiments.

La fusion IMU-GNSS permet ainsi de tirer parti des avantages de chaque capteur :

- L'IMU assure une propagation continue et réactive de l'état,

- Le GNSS joue le rôle de référence corrective, limitant la dérive inertielle.

Cette complémentarité constitue le fondement du recours à un EKF, particulièrement adapté aux systèmes non linéaires et aux dynamiques humaines.

Cette approche marque une évolution par rapport à nos travaux antérieurs, principalement orientés vers la localisation intérieure et fondés sur des combinaisons CNN–Kalman [107]. Le travail présenté ici se concentre spécifiquement sur la localisation extérieure continue, en privilégiant une fusion inertielle–GNSS optimisée et robuste.

3.6.2.1. Prétraitement, calibrage et synchronisation des capteurs

Avant leur intégration dans le processus de fusion, les mesures issues de l'IMU et du GNSS sont prétraitées afin de limiter les erreurs systématiques susceptibles de dégrader l'estimation de l'état.

L'accéléromètre est calibré afin de corriger les biais statiques et les erreurs de gain, tandis que le gyroscope est soumis à une compensation de dérive à taux nul ainsi qu'à une correction des désalignements inter-axes. Ces étapes sont essentielles pour éviter l'amplification des erreurs lors de la phase d'intégration inertielle.

Par ailleurs, compte tenu des fréquences d'échantillonnage différentes des capteurs, une synchronisation temporelle est mise en œuvre. Les mesures GNSS sont interpolées afin de correspondre aux instants de propagation inertielle, garantissant ainsi la cohérence temporelle du processus de fusion.

3.6.2.2. Modélisation de l'état du système

Le vecteur d'état du filtre est conçu pour représenter de manière complète l'état cinématique de l'utilisateur ainsi que certaines grandeurs auxiliaires nécessaires à l'estimation :

$$X_k = \begin{bmatrix} p_k \\ v_k \\ \theta_k \\ m_k \end{bmatrix} \quad (26)$$

Où : $p_k = [p_x, p_y, p_z]^T$: Coordonnées de position selon les axes est, nord et vertical (m)

$v_k = [v_x, v_y, v_z]^T$: Composantes de vitesse selon les mêmes axes est, nord et vertical (m/s)

$\theta_k = [\phi, \theta, \psi]^T$: désigne l'orientation exprimée par les angles d'Euler (roulis, tangage, cap)

$m_k = [m_x, m_y, m_z]^T$: Mesures du champ magnétique terrestre selon les axes du capteur.

Cette modélisation permet d'établir un lien direct entre les mesures physiques des capteurs et l'état interne du filtre.

▪ **Phase de prédiction :**

La phase de prédiction constitue le cœur de la continuité de la localisation. Elle est exécutée à chaque pas de temps IMU et repose exclusivement sur les mesures inertielles, indépendamment de la disponibilité du GNSS. À partir de l'état estimé à l'instant « $k-1$ », les mesures d'accélération « a_k » et de vitesse angulaire « ω_k » sont utilisées pour propager l'état vers l'instant « k » selon un modèle cinématique non linéaire :

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1}, a_k, \omega_k) + w_k \quad (27)$$

Sur le plan fonctionnel, cette propagation inertielle s'effectue en plusieurs étapes successives. Les vitesses angulaires sont d'abord intégrées afin de mettre à jour l'orientation de l'utilisateur. Cette orientation est ensuite utilisée pour projeter les accélérations mesurées dans le repère global, permettant ainsi l'estimation de la vitesse puis de la position par intégration successive.

Ce mécanisme permet de suivre avec précision les mouvements instantanés de l'utilisateur, même en l'absence temporaire de mesures GNSS. Toutefois, l'incertitude associée à cette estimation augmente progressivement, ce qui est reflété par la propagation de la matrice de covariance par Eq.20.

Cette augmentation de l'incertitude traduit la confiance décroissante accordée à l'estimation inertielle à mesure que le temps s'écoule sans correction externe.

▪ **Phase de correction : mise à jour par mesures GNSS :**

Lorsque des mesures GNSS sont disponibles, elles sont utilisées pour corriger l'estimation prédite et limiter la dérive inertielle. Le vecteur de mesure est défini comme :

$$z_k = [p_k^{GNSS} \ v_k^{GNSS} \ m_k^{mag}]^T \quad (28)$$

La phase de correction repose sur le calcul de l'innovation, qui représente l'écart entre la position prédite par le modèle inertiel et la position mesurée par le GNSS. Cette innovation est pondérée par les incertitudes respectives afin de mettre à jour l'état estimé par Eq 24. Le gain de Kalman « K_k » joue un rôle central en modulant l'influence des mesures GNSS. Ainsi, lorsque le signal GNSS est fiable, la correction est significative, tandis qu'en cas de dégradation du signal, le filtre privilégie la continuité inertielle.

3.6.2.3. Adaptation dynamique des covariances

Afin d'améliorer la robustesse du système face aux variations de dynamique de l'utilisateur et aux fluctuations de la qualité du signal GNSS, les matrices de covariance « Q_k » et « R_k » sont adaptées dynamiquement. Cette adaptation permet au filtre d'ajuster automatiquement son niveau de confiance dans chaque source d'information, renforçant ainsi la stabilité de l'estimation dans des environnements complexes.

Dans le contexte de l'assistance à la mobilité des personnes non-voyantes, l'approche de localisation considérée fournit une estimation continue, fiable et exploitable de la position de l'utilisateur. Cette estimation constitue une information centrale pour l'ensemble du système, notamment pour le guidage, l'anticipation des situations à risque et la coordination avec le sous-système d'évitement des obstacles. En assurant une localisation robuste même en conditions GNSS dégradées, la fusion IMU–GNSS via EKF contribue directement à renforcer l'autonomie de l'utilisateur, en lui permettant de se déplacer de manière plus sûre, plus fluide et plus confiante dans des environnements extérieurs complexes.

3.7. Interface auditive et implémentation embarquée

L'efficacité d'un système d'assistance à la navigation pour personnes non-voyantes repose non seulement sur la précision des algorithmes de perception, mais également sur leur intégration matérielle et leur restitution à l'utilisateur. Cette section présente la conception de l'interface auditive intelligente et les choix d'implémentation embarquée permettant de délivrer, en temps réel et de manière compréhensible, les informations critiques générées par les modules de vision et de décision. Nous détaillons le processus de synthèse et de priorisation des messages vocaux, ainsi que la sélection de la plateforme matérielle garantissant l'autonomie, la fiabilité et la portabilité de l'ensemble du système.

3.7.1. Génération du retour audio contextuel pour l'assistance à la navigation

Le retour auditif constitue l'interface homme-machine principale du système, destinée à transmettre à l'utilisateur non-voyant les informations critiques issues des sous-systèmes de perception, d'analyse et de décision. L'objectif est de convertir en temps réel des données multimodales en messages vocaux synthétiques, compréhensibles et adaptés au contexte de navigation. Les informations exploitées proviennent des quatre sous-systèmes fondamentaux : détection d'obstacles, classification comportementale, stratégie d'évitement et localisation. Un module décisionnel central fusionne ces flux et priorise les événements selon leur niveau de risque basé sur la distance, la vitesse relative et la probabilité de collision.

Le module de génération sémantique produit des messages vocaux normalisés selon trois catégories :

- Alertes critiques : « Banc statique à 2 mètres à droite », « Piéton dynamique à gauche ».
- Guidage de navigation : « Tourner à droite », « Danger, faites demi-Tour ».
- Informations contextuelles : « Entrée du restaurant dans 2 mètres », « Arrivée à destination ».

Les messages textuels sont convertis en audio via un moteur TTS (Text-To-Speech) embarqué, fonctionnant hors connexion. Une gestion dynamique des priorités permet d'interrompre ou de différer

les messages non urgents lors de l'apparition d'une situation critique. La diffusion s'effectue via un haut-parleur à conduction osseuse, préservant la perception des sons environnementaux.

3.7.2. Choix et justification de la plateforme embarquée

La mise en œuvre temps réel d'un système d'assistance à la navigation impose le choix d'une plateforme embarquée conciliant puissance de calcul, portabilité, faible consommation et compatibilité avec les périphériques requis. La Raspberry Pi 4 (8 Go RAM) a été retenue comme unité de calcul centrale pour les raisons suivantes :

- Performances adaptées : Architecture ARM Cortex-A72 quad-core à 1,8 GHz, capable d'exécuter simultanément l'inférence des modèles de vision, la fusion décisionnelle et la synthèse vocale avec des latences inférieures à 500 ms.
- Connectivité intégrée : Interfaces USB 3.0 pour la caméra, GPIO pour les capteurs annexes, sortie audio HDMI/3,5 mm, et connectivité Bluetooth/Wi-Fi pour les modules de localisation complémentaires.
- Système d'exploitation : Compatibilité avec Linux (Raspberry Pi OS) permettant l'utilisation de bibliothèques optimisées (TensorFlow Lite, PyTorch Mobile, OpenCV) et la gestion temps réel des processus.
- Consommation maîtrisée : TDP de 7,5 W sous charge maximale, compatible avec une alimentation par batterie lithium-ion pour une autonomie supérieure à 4 heures.
- Écosystème mature : Large communauté de support, disponibilité de shields dédiés et possibilité d'intégration dans un boîtier portable.

Cette plateforme représente un compromis technique pertinent entre capacités de traitement, modularité et contraintes d'intégration physique, tout en permettant une évolution future vers des accélérateurs matériels dédiés en cas de besoins computationnels accrus.

3.8. Conclusion

Ce chapitre a présenté notre méthodologie de système d'assistance à la navigation pour les personnes non-voyantes, structurée autour de quatre sous-systèmes fondamentaux :

- Détection d'obstacles avec YOLOv7-Plus optimisé.
- Classification comportementale via fusion ConvLSTM-DINOv2.

- Stratégie d'évitement basée sur une logique floue.
- Localisation utilisant un filtre de Kalman étendu (EKF) avec fusion de capteurs.

Cette approche intégrée permet de transformer un environnement complexe en informations exploitables pour la navigation autonome. Notre architecture concilie sécurité, orientation et gestion de la charge cognitive, tout en garantissant une interaction naturelle et fiable avec l'utilisateur.

Cette architecture complète pour la navigation extérieure pose cependant les limites de son applicabilité en environnement intérieur, où l'absence de signaux GPS et la complexité des espaces clos nécessitent une approche adaptée.

Le chapitre suivant présentera donc la conception et la réalisation d'un système d'assistance "indoor" spécifique, détaillant les méthodes de localisation en espace confiné, la modélisation des environnements bâtis, et les stratégies de guidage adaptées aux caractéristiques architecturales des milieux intérieurs. Cette extension permettra d'offrir une solution d'assistance complète et continue, de l'espace public extérieur jusqu'aux destinations précises au sein des bâtiments.

Chapitre 4 : Conception d'un système de navigation indoor pour les non-voyants

4.1. Introduction

Le Chapitre 3 de cette thèse a été consacré à la conception méthodologique et algorithmique d'un système intelligent d'assistance à la navigation destiné aux personnes non-voyantes. Il a détaillé les principes théoriques, les modèles algorithmiques et les stratégies de traitement associés aux différents sous-systèmes du dispositif, notamment la perception visuelle, l'analyse spatio-temporelle, la localisation et la génération de retours auditifs. Ces développements ont été étudiés principalement dans un cadre outdoor, à travers des scénarios simulés et des validations modulaires permettant d'évaluer séparément les performances de chaque composant. Cette approche par simulation et validation indépendante des sous-systèmes a permis de démontrer la pertinence conceptuelle des méthodes proposées, ainsi que leur potentiel en termes de précision, de réactivité et de robustesse. Toutefois, une telle validation reste insuffisante pour apprécier les contraintes réelles d'un déploiement embarqué, en particulier lorsque le système est destiné à un usage quotidien par des utilisateurs non-voyants.

En effet, les environnements indoor introduisent des problématiques spécifiques qui ne peuvent être pleinement appréhendées par la simulation seule. L'absence de signaux GNSS, la proximité constante des obstacles, la diversité du mobilier, la variabilité des conditions acoustiques et la nécessité d'une interaction rapide et intuitive imposent une intégration étroite des sous-systèmes ainsi qu'une optimisation rigoureuse des algorithmes pour l'exécution en temps réel sur une plateforme embarquée à ressources limitées.

Dans ce contexte, le présent chapitre vise à transposer les fondements méthodologiques établis au Chapitre 3 vers une réalisation concrète, en se focalisant sur la conception et l'implémentation d'un système de navigation indoor opérationnel. Contrairement au chapitre précédent, l'accent est ici mis sur l'architecture matérielle et logicielle, le choix des capteurs, la plateforme embarquée, ainsi que les mécanismes d'intégration entre perception, localisation et retour utilisateur.

Ce chapitre présente ainsi la réalisation complète d'un prototype fonctionnel, conçu selon une démarche centrée utilisateur, et optimisé pour un usage réel par des personnes non-voyantes. Il décrit successivement la conception du dispositif, les choix technologiques retenus, l'adaptation des algorithmes aux contraintes indoor, ainsi que les stratégies de fusion des informations issues des différents sous-systèmes. Enfin, une évaluation expérimentale en environnement intérieur contrôlé, menée auprès d'utilisateurs non-voyants, permet de mesurer les performances du système en termes d'efficacité de navigation, de sécurité et de charge cognitive.

En résumé, ce chapitre constitue une étape essentielle de validation expérimentale du travail de thèse, en assurant la transition entre une approche algorithmique fondée sur la simulation et une implémentation embarquée évaluée en conditions proches de l'usage réel. Il démontre la faisabilité, la

cohérence et l'intérêt pratique du système proposé pour l'assistance à la navigation indoor des personnes non-voyantes.

4.2. Architecture globale du système

L'architecture globale du système proposé repose sur une approche intégrée combinant perception multimodale, traitement embarqué, fusion de capteurs et interaction audio, afin d'assurer une assistance efficace à la navigation indoor des personnes non-voyantes. Le système a été conçu, implémenté et validé sous forme de prototype fonctionnel, en tenant compte des contraintes spécifiques aux environnements intérieurs et aux besoins des utilisateurs. La Figure 4.1 présente une vue d'ensemble de cette architecture, illustrant les différents modules matériels et logiciels ainsi que les flux d'information entre eux.

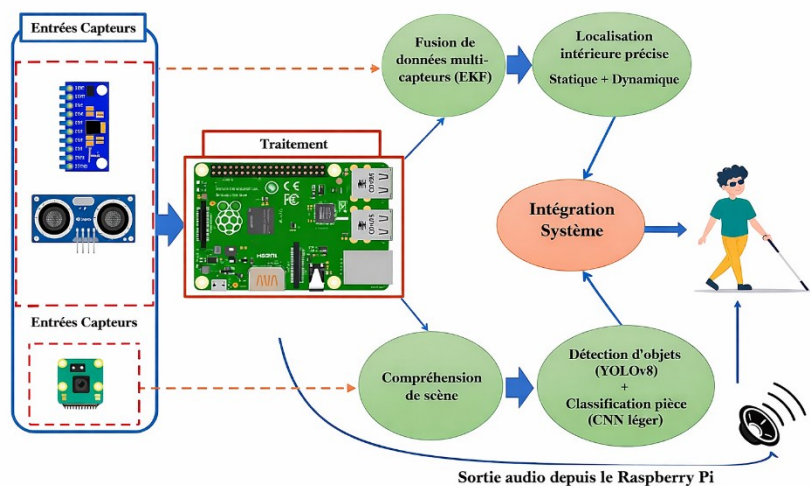


Figure 4.1. Architecture générale du système de navigation indoor

4.2.1. Principes architecturaux et structuration fonctionnelle du système

La conception du dispositif s'inscrit dans une démarche de design centré utilisateur, fondée sur les principes du design inclusif et universel, visant à garantir l'accessibilité fonctionnelle du système pour les personnes non-voyantes tout en préservant son utilisabilité dans des contextes d'usage quotidiens. Cette démarche a conduit à considérer la portabilité, l'accessibilité économique et la facilité de prise en main comme des contraintes de conception structurantes, conditionnant à la fois l'acceptabilité du dispositif et son intégration durable dans les pratiques de mobilité. Les besoins utilisateurs ont été identifiés à partir de retours qualitatifs recueillis lors d'entretiens informels menés auprès des membres de l'association Israr à Annaba (Algérie). Ces échanges ont mis en évidence trois exigences majeures. La première concerne la nécessité d'un retour auditif à très faible latence, afin de préserver la continuité

perceptive entre l'évolution de l'environnement et la prise de décision de l'utilisateur, élément critique pour la sécurité lors des déplacements. La seconde porte sur les contraintes physiques du système, lequel doit présenter un encombrement réduit, une masse minimale et une discrétion suffisante pour ne pas altérer le confort, la liberté de mouvement ni l'acceptation sociale lors d'un port prolongé. La troisième exigence vise la maîtrise de la charge cognitive, impliquant que les informations restituées soient pertinentes, hiérarchisées et immédiatement interprétables, afin d'éviter toute surcharge susceptible de nuire à la réactivité de l'utilisateur.

Sur la base de ces contraintes, le système a été conçu comme un dispositif portable porté sur le corps, pouvant être fixé de manière non intrusive à un vêtement ou à un accessoire personnel. Ce choix de conception favorise une interaction naturelle avec le système et limite les adaptations comportementales requises de la part de l'utilisateur. L'interface audio a été spécifiquement pensée pour conserver une intelligibilité élevée dans des environnements acoustiquement complexes, tout en s'intégrant de manière complémentaire aux capacités d'écholocalisation naturelle et aux dispositifs d'assistance traditionnels, tels que la canne blanche. Cette complémentarité vise à enrichir la perception spatiale sans interférer avec les stratégies de navigation déjà acquises, renforçant ainsi la sécurité et l'autonomie de l'utilisateur.

4.2.2. Architecture matérielle embarquée

L'architecture matérielle embarquée du système repose sur l'utilisation d'une plateforme Raspberry Pi 4 Model B équipée de 8 Go de RAM, qui constitue le noyau de traitement principal. Le choix de cette plateforme résulte d'une analyse multicritère visant à concilier : capacité de calcul, efficacité énergétique, coût de mise en œuvre et accessibilité matérielle, conformément aux contraintes propres à un dispositif d'assistance portable destiné à un usage quotidien.

Contrairement aux plateformes de calcul dédiées à l'intelligence artificielle en périphérie (AI edge modules), qui offrent des performances de calcul supérieures mais s'accompagnent généralement d'une consommation énergétique élevée, d'une dissipation thermique accrue et d'un coût matériel important, le Raspberry Pi 4 présente un compromis particulièrement adapté au contexte d'un système embarqué porté sur le corps. Il offre une puissance de calcul suffisante pour l'exécution de modèles de deep learning optimisés, notamment grâce à l'utilisation de TensorFlow Lite et de techniques de quantification des réseaux neuronaux, tout en maintenant un faible encombrement, une consommation maîtrisée et un profil thermique compatible avec une utilisation prolongée.

Comme illustré à la Figure 4.1, la plateforme de calcul est interfacée avec un ensemble de capteurs complémentaires, assurant une perception multimodale de l'environnement. Une caméra Pi Camera v1.3 est dédiée à l'acquisition d'images en temps réel, fournissant les données visuelles nécessaires à la détection d'obstacles et à l'analyse du contexte spatial immédiat. Un capteur inertiel MPU9250 est intégré afin de mesurer les accélérations et les vitesses angulaires, permettant le suivi du mouvement et

de l'orientation de l'utilisateur. En complément, un capteur ultrasonique HC-SR04 est employé pour l'estimation directe de la distance des obstacles proches, renforçant la fiabilité de la perception dans des situations où la vision seule peut s'avérer insuffisante.

L'ensemble du système est alimenté par une batterie portable offrant une autonomie d'environ six heures, répondant aux exigences de mobilité et d'usage prolongé. Les composants électroniques sont intégrés dans un boîtier léger, permettant une fixation discrète et ergonomique sur le corps de l'utilisateur. Le traitement embarqué des données repose sur une chaîne logicielle optimisée, dans laquelle l'inférence des modèles de perception est réalisée via TensorFlow Lite, garantissant des temps de réponse compatibles avec les contraintes de navigation assistée en temps réel, tout en préservant la stabilité et la réactivité du système.

4.2.3. Architecture logicielle et organisation des flux fonctionnels

L'architecture logicielle du système d'assistance est conçue selon une organisation modulaire et distribuée, afin de satisfaire les exigences de traitement en temps réel, de fiabilité et de continuité opérationnelle propres aux dispositifs de navigation indoor destinés aux personnes non-voyantes. Cette organisation vise à assurer une gestion efficace des flux de données issus de sources multiples, tout en maintenant une cohérence temporelle entre les différents traitements nécessaires à la génération d'une assistance pertinente.

Le traitement des données visuelles constitue l'un des axes centraux de l'architecture logicielle. Les images acquises en temps réel par la caméra embarquée sont analysées à l'aide de modèles de deep learning optimisés pour une exécution sur plateforme embarquée. La détection des obstacles repose sur l'utilisation du modèle YOLOv8s [108], sélectionné pour sa capacité à fournir une localisation et une identification rapides des objets pertinents pour la navigation, tout en maintenant une latence compatible avec les contraintes d'un système portable. Contrairement au chapitre 3, qui présentait une architecture orientée vers des environnements extérieurs plus généraux, le présent chapitre adopte la variante légère YOLOv8s afin de répondre aux contraintes de calcul et de consommation de ressources du prototype indoor embarqué sur Raspberry Pi 4B, tout en garantissant des performances de détection adaptées à une exécution en temps réel. En complément, un réseau MobileNetV2 [109] affiné est employé pour la classification des environnements indoor, permettant d'identifier des contextes structurants tels que les couloirs, escaliers ou espaces fonctionnels. L'association de ces traitements visuels permet d'obtenir une description sémantique de l'environnement immédiat, intégrant à la fois la présence d'obstacles et la nature de l'espace parcouru.

L'estimation de la position et du mouvement de l'utilisateur en environnement indoor, caractérisé par l'absence de signal GNSS, repose sur la fusion des données issues du capteur inertiel MPU9250 et du capteur ultrasonique. Les mesures d'accélération et de vitesse angulaire sont exploitées pour modéliser

la dynamique du déplacement, tandis qu'un filtre de Kalman étendu est utilisé afin d'estimer l'état du système en tenant compte des incertitudes associées aux capteurs. Cette approche permet de limiter l'accumulation des erreurs liées au bruit et à la dérive des mesures inertielles. Les informations de distance fournies par le capteur ultrasonique sont intégrées afin de renforcer la perception spatiale à courte portée, en particulier dans les environnements confinés ou présentant une forte densité d'obstacles.

Les informations issues des modules de perception et de localisation sont ensuite transmises au module d'interaction utilisateur, chargé de la génération du retour vocal. Ce module transforme les informations spatio-contextuelles en messages synthétiques délivrés de manière progressive et contextualisée. La conception du retour vocal vise à transmettre les informations essentielles à la prise de décision de l'utilisateur, tout en minimisant la charge cognitive et en préservant la perception des sons ambiants nécessaires à l'orientation.

L'ensemble du système logiciel fonctionne selon une gestion asynchrone des traitements, chaque composant étant exécuté à une fréquence adaptée à la dynamique des données qu'il traite. Cette organisation permet de réduire la latence globale, d'améliorer la stabilité des estimations et de garantir une assistance continue, y compris en présence de variations des conditions environnementales ou de la qualité des mesures. Par ailleurs, cette structuration favorise l'optimisation progressive des algorithmes et facilite l'évolution du système vers des configurations plus complexes, sans remise en cause de l'architecture logicielle globale.

4.3.Implémentation fonctionnelle des sous-systèmes

Cette section présente l'implémentation fonctionnelle des sous-systèmes constitutifs du dispositif d'assistance à la navigation indoor proposé. Elle vise à expliciter la manière dont les modules de perception, de localisation et d'interaction utilisateur sont concrètement mis en œuvre et articulés afin de garantir un fonctionnement cohérent, réactif et fiable dans des environnements intérieurs complexes.

L'accent est porté sur la traduction opérationnelle des choix méthodologiques exposés dans les sections précédentes, en tenant compte des contraintes propres à un système embarqué portable, notamment en termes de latence, de ressources computationnelles et de continuité du service. Chaque sous-système est conçu pour remplir une fonction bien définie, tout en contribuant à une chaîne de traitement globale assurant l'interprétation de l'environnement, l'estimation du mouvement de l'utilisateur et la génération d'un retour informatif exploitable en temps réel.

L'organisation présentée dans cette section met en évidence la complémentarité des traitements visuels, inertiels et proximétriques, ainsi que les mécanismes assurant leur coordination temporelle. Cette

structuration fonctionnelle constitue un élément clé pour garantir une assistance stable, intelligible et adaptée aux besoins spécifiques des personnes non-voyantes lors de leurs déplacements en environnement indoor.

4.3.1. Classification contextuelle des environnements indoor

La classification contextuelle des environnements indoor constitue une fonction centrale du module perceptif du système d'assistance à la navigation proposé. Elle vise à identifier la nature de l'espace dans lequel évolue l'utilisateur afin d'introduire une compréhension sémantique globale de la scène observée. Cette information contextuelle joue un rôle déterminant dans l'interprétation des éléments visuels et dans l'adaptation des mécanismes d'assistance aux contraintes propres à chaque type d'environnement intérieur. Dans un cadre de navigation pour personnes non-voyantes, la simple détection d'objets isolés ne suffit pas à fournir une représentation exploitable de l'espace. La reconnaissance du contexte spatial permet de structurer l'analyse visuelle, d'anticiper les configurations potentielles de l'environnement et d'améliorer la pertinence des informations transmises à l'utilisateur, tout en limitant la surcharge cognitive.

La reconnaissance d'environnements indoor repose sur l'analyse de motifs visuels complexes, intégrant des indices architecturaux, des structures spatiales et des éléments fonctionnels. Les réseaux de neurones convolutifs (CNN) [110] constituent une approche de référence pour ce type de tâche, en raison de leur capacité à apprendre automatiquement des représentations hiérarchiques à partir d'images. Les couches convolutionnelles successives permettent d'extraire progressivement des caractéristiques visuelles allant des contours et textures élémentaires jusqu'à des configurations globales caractéristiques d'un type de pièce. Par rapport aux méthodes reposant sur des descripteurs visuels conçus manuellement, les CNN offrent une meilleure robustesse face aux variations d'éclairage, d'angle de vue et d'encombrement, conditions fréquemment rencontrées dans les environnements indoor réels. Cette propriété est essentielle pour garantir une classification fiable dans un système embarqué utilisé en situation de mobilité.

Bien que de nombreuses architectures CNN profondes présentent des performances élevées en classification de scènes, leur déploiement sur une plateforme embarquée portable impose des contraintes strictes en termes de complexité computationnelle, de consommation énergétique et de latence [111]. Dans ce contexte, l'architecture MobileNetV2 a été retenue comme solution adaptée aux exigences du système.

MobileNetV2 est une architecture de CNN optimisée pour les dispositifs à ressources limitées. Elle s'appuie sur des convolutions séparables en profondeur et des blocs à goulot inversé avec des connexions résiduelles, permettant de réduire significativement le nombre de paramètres et le coût de

calcul tout en maintenant une capacité de représentation suffisante pour la classification de scènes. Ce compromis garantit une exécution efficace sur la plateforme embarquée sans dégradation notable des performances de reconnaissance. L'architecture détaillée de l'approche suggérée est illustrée à la Figure 4.2.

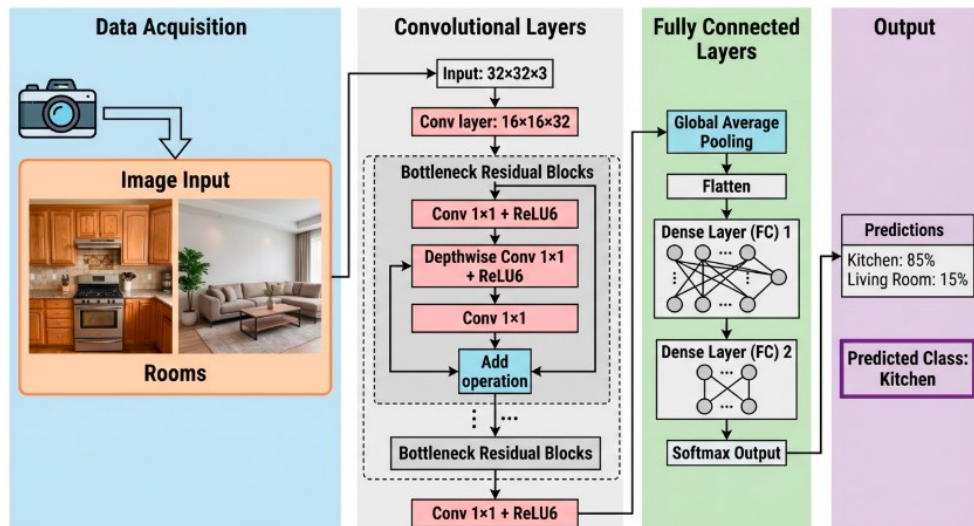


Figure 4.2. Diagramme de l'architecture de classification indoor

L'apprentissage du modèle s'appuie sur un sous-ensemble de la base de données MIT Indoor Scenes [112], reconnu pour sa diversité et sa représentativité des environnements intérieurs. Un total de 7 370 images a été extrait de cette base et réparti entre les ensembles d'apprentissage, de validation et de test. Afin d'assurer une adéquation directe aux besoins de la navigation assistée, 11 classes d'environnements indoor ont été retenues, correspondant à des espaces couramment rencontrés dans la vie quotidienne et présentant des contraintes de déplacement distinctes, telles que les couloirs, les cuisines, les escaliers et les espaces de séjour.

Des techniques d'augmentation de données ont été appliquées lors de l'entraînement afin d'améliorer la robustesse du modèle face aux variations d'illumination, de perspective et de composition visuelle. Cette stratégie contribue à renforcer la capacité de généralisation du réseau lors de son déploiement en conditions réelles.

Lors du fonctionnement du système, les images acquises par la caméra embarquée sont prétraitées, puis analysées par le réseau MobileNetV2 afin de produire une estimation du contexte environnemental. L'inférence est réalisée à l'aide de TensorFlow Lite, permettant une exécution optimisée compatible avec les contraintes matérielles de la plateforme. La classification est effectuée à une fréquence d'environ une image toutes les 1,5 secondes, assurant un compromis entre réactivité du système et maîtrise de la charge computationnelle. Pour chaque image traitée, le réseau fournit une distribution de

probabilités associée aux différentes classes. La classe présentant la probabilité la plus élevée est retenue comme représentation du contexte dominant et transmise aux modules perceptifs suivants.

La classification contextuelle fournit ainsi une interprétation sémantique structurante de la scène indoor. Cette information est exploitée conjointement avec les résultats de la détection d'obstacles afin d'affiner l'analyse visuelle et d'en améliorer la cohérence. En effet, la signification fonctionnelle d'un objet dépend fortement de l'environnement dans lequel il est observé.

En intégrant explicitement le contexte environnemental, le système est en mesure de produire une représentation visuelle plus claire et mieux structurée de l'espace indoor, contribuant à une assistance à la navigation plus pertinente, plus fiable et mieux adaptée aux besoins des utilisateurs non-voyants.

4.3.2. Détection des obstacles

La détection des obstacles intervient comme la seconde fonction perceptive du système, après l'identification contextuelle de l'environnement indoor. Alors que la classification des scènes fournit une compréhension globale du lieu dans lequel évolue l'utilisateur, la détection d'objets vise à identifier, dans ce contexte, les éléments physiques susceptibles d'interférer avec le déplacement. Cette articulation entre information contextuelle et perception locale permet d'orienter la détection vers les obstacles réellement pertinents et d'en améliorer l'interprétabilité pour l'utilisateur non-voyant.

Dans ce cadre, la détection est assurée par un réseau profond de type YOLO (You Only Look Once), dont le principe repose sur une formulation unifiée du problème de détection. Contrairement aux approches séquentielles, YOLO traite chaque image comme une entrée globale et apprend simultanément à prédire la présence des objets, leur catégorie et leur position spatiale. Cette formulation permet une inférence directe et continue, compatible avec les contraintes temporelles d'un système d'assistance à la navigation fonctionnant en conditions réelles.

Après l'étape de classification contextuelle, chaque image acquise par la caméra est transmise au module YOLO, qui procède à son analyse en une seule passe du réseau. L'image est d'abord transformée en une représentation hiérarchique de caractéristiques visuelles au moyen d'un backbone convolutionnel, chargé d'extraire les structures pertinentes à différentes échelles spatiales. Ces caractéristiques sont ensuite propagées vers les couches de détection, où le réseau prédit, pour chaque région candidate de l'image, des boîtes englobantes associées à des probabilités de présence d'objet et à des scores de classification.

Dans le système proposé, la version YOLOv8s a été retenue en raison de sa conception architecturale récente, spécifiquement orientée vers une meilleure efficacité sur des plateformes embarquées [113]. YOLOv8s adopte un backbone de type C2f, favorisant une meilleure propagation des gradients et une extraction plus stable des caractéristiques multi-échelles, tout en limitant la complexité du modèle. De

plus, l'abandon des ancres prédéfinies au profit d'une formulation anchor-free permet une régression plus directe des boîtes englobantes, réduisant les erreurs de localisation et simplifiant le processus d'apprentissage.

Les Heads de détection de YOLOv8s sont structurés de manière à séparer implicitement les tâches de classification et de régression spatiale, ce qui améliore la précision de la détection sans introduire de surcharge computationnelle significative. Cette organisation favorise une meilleure adaptation du modèle aux variations d'échelle et de forme des obstacles rencontrés en environnement indoor, tels que les chaises, portes, escaliers ou tables.

Le modèle a été affiné sur un sous-ensemble de la base de données indoor Open Images [114], comprenant 13 928 images annotées réparties en treize catégories d'objets directement liées à la mobilité des personnes non-voyantes. Cette étape d'adaptation permet d'aligner les capacités du détecteur sur les caractéristiques visuelles propres aux environnements intérieurs et aux obstacles qu'ils contiennent. Une optimisation du modèle a ensuite été réalisée afin d'assurer une exécution efficace sur la plateforme embarquée. Les expérimentations ont été conduites avec YOLOv8s, en utilisant une résolution d'entrée de 640×640 pixels, sur un Raspberry Pi 4B exécutant les inférences uniquement sur le CPU, sans accélération matérielle. Le modèle, converti au format TensorFlow Lite avec quantification entière (INT8), a été évalué dans des conditions de fonctionnement réelles correspondant à un usage indoor, avec un flux vidéo issu de la caméra embarquée. Dans cette configuration, le système atteint un débit d'inférence compris entre 8 et 11 images par seconde, avec une latence moyenne de 90 à 125 ms par image et une latence maximale d'environ 140 ms, garantissant une exécution compatible avec les exigences de la navigation assistée en temps réel. La figure 4.3 illustre le processus de détection proposé.

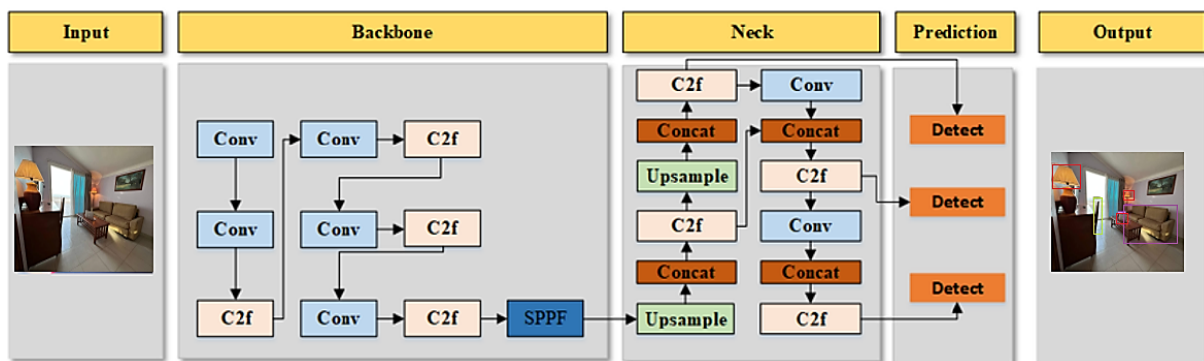


Figure 4.3. Architecture du processus de détection d'obstacle pour environnement indoor

Les objets détectés sont ensuite mis en relation avec les résultats de la classification contextuelle de la scène. Cette association permet de renforcer la cohérence sémantique de la perception, en interprétant chaque obstacle au regard du type d'environnement reconnu. Par exemple, la détection d'une table n'a pas la même signification fonctionnelle selon qu'elle se situe dans une cuisine, une salle de lecture ou un couloir, ce qui influe directement sur la priorité et la formulation du retour utilisateur.

Toutefois, la détection visuelle, bien qu'elle fournisse une information sémantique et une localisation relative dans le plan image, ne permet pas à elle seule de déterminer avec précision la position spatiale de l'obstacle par rapport à l'utilisateur. Afin de transformer ces résultats perceptifs en indications exploitables pour la navigation, il est nécessaire d'estimer la position réelle de l'obstacle dans l'espace et de la synchroniser avec le mouvement de l'utilisateur. Cette problématique conduit naturellement à la mise en place d'un sous-système de localisation et d'estimation spatiale, présenté dans la section suivante, qui vise à fournir une localisation métrique cohérente et temporellement stable des obstacles détectés.

4.3.3. Localisation et estimation du mouvement en environnement indoor

L'estimation précise du mouvement de l'utilisateur et de la position des obstacles est un élément déterminant pour la sécurité et l'efficacité d'un système de navigation indoor destiné aux personnes non-voyantes. Dans ce contexte, après la classification contextuelle des environnements et la détection des objets, il est essentiel de localiser les obstacles de manière métrique pour permettre des décisions de navigation fiables et prévenir les collisions. La solution adoptée repose sur une fusion de capteurs inertiels et ultrasoniques traitée via un Filtre de Kalman Étendu (EKF), comme illustré à la Figure 4.4.

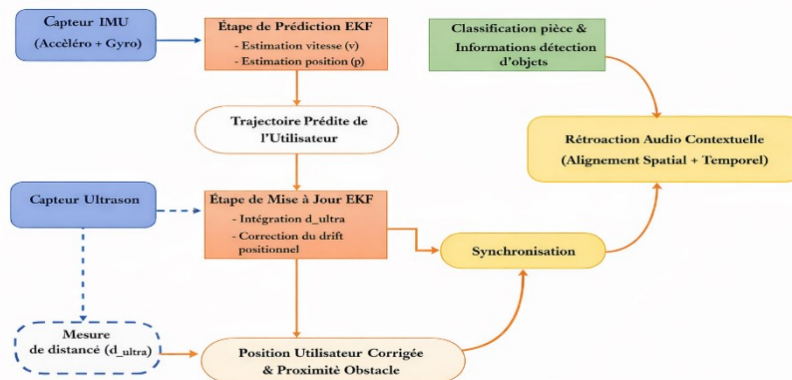


Figure 4.4. Diagramme du processus de localisation

Le prototype utilise un capteur inertiel MPU9250 (accéléromètre et gyroscope) pour estimer la vitesse et l'orientation de l'utilisateur en continu. Ce choix repose sur la capacité de l'IMU à fournir des mesures haute fréquence (50 Hz) permettant de suivre les mouvements dynamiques avec une latence minimale. En parallèle, un capteur ultrasonique HC-SR04 est intégré pour fournir des mesures absolues de distance aux obstacles, à une fréquence de 10 Hz. Ce capteur compense la dérive inévitable de l'intégration inertielle, garantissant une correction externe fiable.

L'Extended Kalman Filter est utilisé par la nécessité de gérer à la fois les non-linéarités de la dynamique de déplacement de l'utilisateur et les incertitudes différentes des capteurs. Il permet d'intégrer de

manière optimale des mesures hétérogènes à fréquence différente et de produire une estimation de l'état du système (position et orientation) avec un compromis calculatoire adapté aux plateformes embarquées. La fusion de l'IMU et de l'ultrasonique au sein de l'EKF offre une localisation robuste, capable de résister aux perturbations courantes des environnements intérieurs (interférences magnétiques, obstacles variés, signal GPS indisponible).

Le filtre fonctionne en deux étapes principales, prédiction et mise à jour, exécutées de manière récursive à chaque intervalle d'échantillonnage :

- Etape de prédiction :

À chaque intervalle « Δt » la position « p_t » et la vitesse « v_t » de l'utilisateur sont estimées à partir des mesures d'accélération « a_t » :

$$p_{t+1} = p_t + v_t \Delta t + \frac{1}{2} a_t (\Delta t)^2 \quad (29)$$

$$v_{t+1} = v_t + a_t \Delta t \quad (30)$$

Cette étape projette l'état futur du système en se basant sur les mouvements mesurés par l'IMU, tout en mettant à jour la covariance associée à l'estimation pour tenir compte des incertitudes du modèle.

- Etape de mise à jour

Les mesures de distance « z_t » fournies par le capteur ultrasonique corrigent la prédiction afin de réduire la dérive, particulièrement dans la direction du déplacement. L'équation de mise à jour s'écrit :

$$z_t = h(p_t) + v_t \quad (31)$$

Où « $h(p_t)$ » représente la distance anticipée à un obstacle en fonction de la position estimée, et « v_t » le bruit gaussien de mesure. Le gain de Kalman permet de pondérer de manière optimale la confiance accordée à la prédiction IMU et aux mesures ultrasoniques.

L'estimation métrique obtenue par l'EKF est synchronisée avec les objets détectés par YOLO et avec la classification contextuelle des environnements. Cette intégration permet de positionner précisément chaque obstacle dans l'espace et d'adapter la trajectoire de l'utilisateur en conséquence. Par exemple, une table détectée dans une cuisine ou un salon sera localisée avec exactitude, permettant d'éviter les collisions et de planifier un chemin sûr dans des espaces restreints ou encombrés.

Cette approche garantit une localisation fiable des obstacles et contribue directement à la prévention des collisions, assurant ainsi la sécurité et la confiance de l'utilisateur dans son environnement.

L'information spatiale ainsi obtenue constitue la base nécessaire pour la génération d'indications vocales contextuelles, étape suivante du système de navigation, qui transforme ces données en guidance exploitables et intuitives pour l'utilisateur.

4.3.4. Génération d'indications vocales contextuelles

La restitution des informations perceptives constitue la dernière étape du traitement sensoriel, assurant une navigation sûre et autonome pour les utilisateurs dans des environnements intérieurs complexes. Dans ce dispositif, le retour d'information est transmis via un haut-parleur positionné sur l'épaule, permettant à l'utilisateur de conserver la perception des sons ambiants, essentiels à l'orientation spatiale et à la détection naturelle des obstacles.

Pour limiter la surcharge cognitive et améliorer la lisibilité des informations transmises, une stratégie hiérarchisée de filtrage et de priorisation a été mise en place :

- Annonce des pièces : les informations relatives à la pièce sont communiquées uniquement lors d'un changement d'environnement, afin d'éviter toute redondance inutile.
- Annonce des obstacles : les messages sont générés uniquement pour les nouvelles détections ou les obstacles critiques, tandis que les objets déjà identifiés ne sont pas répétés.
- Messages concis et explicites : la formulation des retours est calibrée pour être immédiatement compréhensible et exploitable, par exemple : « Vous êtes dans le bureau, table détectée à 2 mètres ».

Ces messages intègrent les informations issues de la classification contextuelle des pièces et de la détection des obstacles, assurant que le retour vocal est cohérent avec l'état réel de l'environnement. Cette combinaison permet à l'utilisateur d'avoir une perception spatio-contextuelle complète, améliorant la réactivité et la sécurité lors de la navigation. L'approche adoptée garantit que l'utilisateur reçoit uniquement les informations essentielles pour le déplacement et l'évitement des obstacles, tout en préservant sa capacité à percevoir l'environnement sonore.

4.4. Conception et réalisation du système

Le prototype final repose sur une intégration matérielle et logicielle optimisée, assurant la combinaison efficace de perception, localisation et interaction utilisateur dans un flux temps réel. Le système est conçu pour un port confortable autour du ventre, fixé par une ceinture ajustable, garantissant une stabilité mécanique et un accès facile aux composants pour l'utilisateur. Cette configuration permet également de maintenir les capteurs dans des positions optimales pour la collecte de données fiables tout en minimisant l'encombrement et la fatigue lors d'une utilisation prolongée.

Les capteurs utilisés dans le système sont positionnés de manière à optimiser la précision des mesures et la sécurité de l'utilisateur :

- Caméra Pi v1.3 : montée sur la face avant de la boîte centrale, elle capture des images pour la classification des pièces et la détection des obstacles, avec un angle de vision adapté pour couvrir la trajectoire immédiate de l'utilisateur.
- IMU MPU9250 : positionnée sur le haut de la boîte centrale portée autour du ventre. Cette position assure une stabilité et une cohérence des mesures d'accélération et de rotation, réduisant l'impact des mouvements du corps sur les données.
- Capteur ultrasonique HC-SR04 : monté dans la main de l'utilisateur pour scanner activement les obstacles proches dans la direction de déplacement, ce qui améliore la précision de l'estimation de la distance et la détection des obstacles mobiles ou partiellement masqués.
- Batterie Power Bank : fournit autonomie et portabilité, permettant un fonctionnement prolongé sans compromettre le confort.
- Raspberry Pi 4B (64-bit) : unité centrale de traitement, exécutant les modèles de deep learning (MobileNetV2 pour la classification et YOLOv8s pour la détection) ainsi que le filtre de Kalman étendu (EKF) pour la localisation.

Pour l'implémentation embarquée, les modèles de classification et de détection ont été convertis en TensorFlow Lite (TFLite) afin de :

- Permettre une exécution optimisée sur le Raspberry Pi 4, compatible avec le traitement vidéo en temps réel.
- Réduire la charge mémoire et le temps de calcul, tout en conservant la précision des modèles.
- Garantir la scalabilité pour intégrer de futurs capteurs ou modèles d'analyse plus avancés sans modifier l'architecture matérielle centrale.

Le flux de traitement complet est ainsi organisé :

- La caméra capture l'environnement pour la classification des pièces et la détection des obstacles.
- L'IMU et le capteur ultrasonique fournissent des mesures fusionnées par l'EKF pour estimer précisément la position, l'orientation et la distance des obstacles, garantissant une localisation fiable et l'évitement des collisions.
- Les informations fusionnées sont transmises au module audio, générant des messages vocaux concis et hiérarchisés, réduisant la charge cognitive et facilitant la navigation sécurisée.

La Figure 4.5 illustre l'ensemble du système monté sur le Raspberry Pi avec la disposition des capteurs et la circulation des données entre perception, localisation et retour utilisateur. Cette conception intégrée

constitue une solution portable, fiable et adaptée aux environnements intérieurs, prête pour les tests expérimentaux et l'évaluation des performances du dispositif.

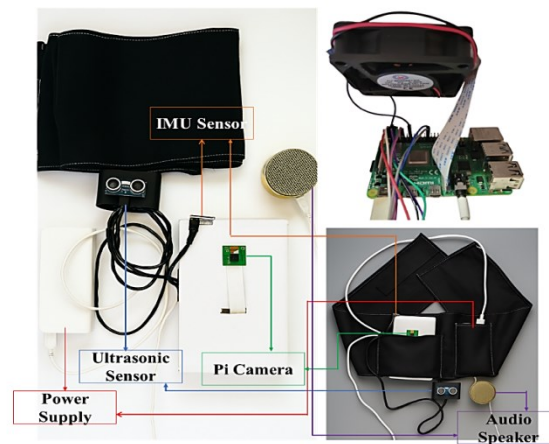


Figure 4.5. Le prototype du système de navigation indoor

4.5. Protocole expérimental et évaluation du système

Dans une perspective d'évaluation en électronique médicale appliquée à l'assistance à la mobilité, le prototype développé a fait l'objet d'une étude expérimentale visant à analyser sa cohérence fonctionnelle, sa pertinence informationnelle et son adéquation avec les besoins d'utilisateurs non-voyants en environnement intérieur. L'évaluation a été conduite dans un cadre structuré, reproduisant des situations réalistes de navigation indoor, avec une intégration complète des modules de classification contextuelle, de détection d'obstacles, de localisation par fusion de capteurs et de restitution vocale.

L'ensemble des procédures a été réalisé dans le respect des principes éthiques internationaux relatifs à la recherche impliquant des personnes humaines, conformément aux recommandations de la Déclaration d'Helsinki [115]. Les participants ont été informés des objectifs du dispositif, de la nature expérimentale de l'évaluation et des modalités de participation. Les échanges et données recueillies ont été traités de manière anonyme afin de garantir la confidentialité et la liberté d'expression.

Douze personnes non-voyantes issues de l'Association ISRAR (Annaba, Algérie) ont pris part à cette étude, en présence du fondateur de l'association. Les profils des participants couvraient une diversité d'âges, de pathologies visuelles et de niveaux de familiarité technologique, permettant une appréciation représentative du système dans un contexte d'usage potentiel. Le tableau 4 présente une synthèse anonymisée des profils.

Table 4. Résumé des profils des participants

ID Participant	Age	Sexe	Type de déficience visuelle	Outils d'assistance utilisés	Familiarité technologique
----------------	-----	------	-----------------------------	------------------------------	---------------------------

P01	25	Homme	Rétinite pigmentaire	Canne blanche	Première utilisation d'un système ETA
P02	33	Femme	Dégénérescence maculaire	Canne blanche, Application smartphone	Utilise quotidiennement un lecteur d'écran
P03	41	Homme	Cécité congénitale	Canne blanche	Première utilisation d'un système ETA
P04	28	Femme	Rétinopathie du prématuré	Application smartphone	Utilisatrice régulière de smartphone
P05	38	Homme	Glaucome	Canne blanche	Déplacement autonome
P06	52	Homme	Rétinopathie diabétique	Application smartphone	À l'aise avec les commandes vocales
P07	35	Femme	Atrophie du nerf optique	Canne blanche, Application smartphone	Préfère les retours tactiles
P08	46	Homme	Dégénérescence maculaire liée à l'âge	Canne blanche	À l'aise avec les dispositifs technologiques
P09	29	Femme	Cataracte congénitale	Application smartphone	Préfère un guidage vocal
P10	50	Homme	Rétinopathie diabétique	Canne blanche, Application smartphone	Réaction plus lente liée à l'âge
P11	26	Femme	Cécité congénitale	Canne blanche, Application smartphone	Passionné de technologie
P12	32	Homme	Rétinite pigmentaire	Application smartphone	Utilise des assistants vocaux

Étant donné que les participants n'étaient pas familiers avec les systèmes de navigation assistée basés sur l'intelligence artificielle, une session introductive structurée a été organisée avant toute mise en situation. Cette phase comprenait une présentation du concept de système électronique d'assistance à la mobilité, une explication détaillée des différents modules fonctionnels (classification contextuelle des environnements, détection d'objets, localisation par fusion de capteurs et génération de messages vocaux), ainsi que des démonstrations pratiques illustrant le fonctionnement du prototype en temps réel. L'objectif était d'assurer une compréhension claire du rôle de chaque composant et de la logique globale du système.

Des scénarios indoor typiques ont ensuite été simulés à l'aide du prototype fonctionnel afin d'illustrer le comportement du dispositif dans différentes configurations spatiales. Parmi les situations présentées figuraient l'entrée dans une nouvelle pièce (par exemple le passage d'un couloir vers un bureau), la détection d'un obstacle frontal à courte distance, la présence d'un obstacle latéral, la coexistence de plusieurs objets dans le champ visuel, ainsi qu'un espace dégagé sans danger immédiat. Pour chaque

scénario, le système était activé et les messages vocaux correspondants étaient générés conformément aux règles définies dans la stratégie d'assistance.

À l'issue de chaque démonstration, les participants ont été invités à évaluer la clarté des messages, la pertinence des informations transmises, le caractère suffisant ou excessif des indications fournies, ainsi que le niveau de confiance qu'ils accorderaient au système dans une situation réelle. Les avis ont été recueillis de manière anonyme afin de garantir une expression libre et non influencée.

L'évaluation du prototype a été conduite selon une approche méthodologique intégrant des instruments normalisés d'analyse de l'utilisabilité, de la charge de travail et de la qualité de l'interaction homme-machine. Cette démarche vise à apprécier non seulement la performance fonctionnelle du système, mais également son adéquation aux exigences ergonomiques propres aux dispositifs d'assistance à la mobilité.

L'utilisabilité globale a été mesurée à l'aide de la « System Usability Scale » (SUS) [116], outil de référence composé de dix items permettant d'obtenir un score quantitatif synthétique. Cet indicateur reflète la facilité d'apprentissage, la cohérence perçue, la simplicité d'utilisation et le niveau de confiance accordé au système. L'emploi de la SUS permet ainsi d'évaluer de manière standardisée l'acceptabilité du dispositif dans une perspective d'usage réel.

La charge de travail subjective a été analysée au moyen du NASA Task Load Index (NASA-TLX) [117], instrument multidimensionnel largement utilisé dans l'évaluation des systèmes interactifs. Cette échelle examine six dimensions complémentaires : demande mentale, demande physique, contrainte temporelle, performance perçue, effort et frustration. Dans le contexte d'un système de navigation assistée pour personnes déficientes visuelles, cette mesure est essentielle afin de vérifier que l'assistance technologique n'introduit pas une surcharge cognitive susceptible de compromettre la sécurité ou le confort d'utilisation.

Par ailleurs, l'analyse ergonomique du système a été conduite en référence aux principes de la norme ISO 9241-110 [118], relatifs à la qualité du dialogue homme-système. Cette norme définit des critères tels que l'adéquation à la tâche, l'auto-descriptivité, la contrôlabilité, la conformité aux attentes de l'utilisateur, la tolérance aux erreurs, l'adaptabilité et la facilité d'apprentissage. L'intégration de ce cadre normatif garantit que la conception ne se limite pas à des performances algorithmiques satisfaisantes, mais qu'elle respecte également les exigences fondamentales d'ergonomie et d'interaction pour un dispositif médical d'assistance à la mobilité.

Cette approche combinée assure une évaluation rigoureuse, multidimensionnelle et conforme aux standards internationaux, permettant d'apprécier la pertinence du système tant sur le plan technique que sur le plan ergonomique et cognitif.

4.6. Conclusion

Ce chapitre a exposé de manière structurée la conception, l'implémentation et la validation méthodologique d'un système embarqué d'assistance à la navigation en environnement indoor dédié aux personnes déficientes visuelles. L'architecture proposée repose sur une intégration cohérente entre perception visuelle, estimation du mouvement et restitution vocale contextuelle, tout en respectant les contraintes inhérentes à un dispositif portable en termes de consommation énergétique, de latence et de capacité de calcul. L'organisation matérielle et logicielle a été pensée pour garantir une exécution en temps réel, une synchronisation fiable des flux de données et une stabilité opérationnelle dans des environnements intérieurs dépourvus de GPS.

Les différents sous-systèmes ont été développés selon une logique fonctionnelle progressive. La classification contextuelle des environnements indoor, assurée par un réseau convolutionnel MobileNetV2 optimisé pour l'embarquer, fournit une information sémantique globale sur la scène observée. Cette information structure l'interprétation des objets détectés par YOLOv8s, dont l'implémentation optimisée permet une détection rapide et précise compatible avec les ressources du Raspberry Pi 4B. La localisation repose sur une fusion de capteurs IMU-ultrason au moyen d'un filtre de Kalman étendu, garantissant une estimation robuste du déplacement et des distances en environnement contraint. L'ensemble de ces traitements converge vers un module de génération d'indications vocales hiérarchisées, conçu pour transmettre une information pertinente tout en maîtrisant la charge cognitive de l'utilisateur.

L'implémentation matérielle et logicielle du prototype final, incluant la conversion des modèles en TensorFlow Lite pour un déploiement efficace sur architecture ARM 64 bits, assure la cohérence entre conception théorique et réalisation pratique. Le protocole expérimental défini, fondé sur des instruments reconnus tels que l'échelle SUS, l'indice de charge NASA-TLX et les principes ergonomiques de la norme ISO 9241-110, établit un cadre rigoureux pour l'évaluation de l'utilisabilité et de l'acceptabilité du système.

Le chapitre suivant est consacré à l'analyse détaillée des performances obtenues et à la discussion des résultats issus des deux approches développées dans ce travail, à savoir la méthodologie de navigation outdoor simulée présentée au chapitre 3 et le système indoor embarqué décrit dans le présent chapitre. Cette analyse permettra d'évaluer la pertinence des choix méthodologiques, de comparer les comportements des deux systèmes et de mettre en évidence les apports scientifiques et applicatifs de cette thèse.

Chapitre 5 : Évaluation expérimentale et analyse des performances des systèmes de navigation assistée

5.1.Introduction

Ce chapitre est consacré à l'évaluation expérimentale et à l'analyse des performances des systèmes de navigation assistée développés dans cette thèse, à savoir l'architecture outdoor simulée et le système indoor embarqué. Il vise à examiner rigoureusement la pertinence des choix méthodologiques adoptés, la robustesse des modèles implémentés et la cohérence globale des solutions proposées au regard des exigences de sécurité, de fiabilité et d'utilisabilité propres aux dispositifs d'assistance à la mobilité. L'approche retenue repose sur une démarche d'évaluation structurée, combinant des indicateurs quantitatifs issus de la vision par ordinateur, du traitement du signal et de la fusion multi-capteurs, ainsi que des critères d'utilisabilité normalisés. Cette double perspective permet d'apprécier à la fois les performances techniques des sous-systèmes développés et leur adéquation aux contraintes réelles d'usage, qu'il s'agisse d'un environnement outdoor simulé ou d'un environnement indoor embarqué.

Ce chapitre a ainsi pour objectif d'analyser les résultats obtenus, d'en discuter la portée scientifique et applicative, et de mettre en évidence les forces et les limites des architectures proposées. Il constitue une étape essentielle dans la validation globale du travail, en assurant la transition vers une discussion approfondie des contributions et des perspectives d'amélioration.

5.2. Évaluation du système de navigation outdoor simulé

L'évaluation du système de navigation outdoor simulé vise à analyser de manière structurée la validité et la robustesse de l'architecture proposée dans un contexte représentatif de la mobilité en milieu extérieur. Bien que cette approche n'ait pas fait l'objet d'une mise en œuvre physique complète, sa validation par simulation permet d'examiner la cohérence algorithmique, la stabilité décisionnelle et les performances des différents modules dans des scénarios réalistes.

L'analyse repose sur une décomposition fonctionnelle du système en sous-systèmes interdépendants, chacun étant évalué selon des critères adaptés à son rôle opérationnel. Cette démarche permet d'identifier la contribution spécifique de chaque module à la sûreté globale du dispositif, notamment en termes de fiabilité de perception, de précision de localisation et de pertinence des décisions d'évitement. L'objectif est d'apprécier la capacité du système à détecter et interpréter correctement son environnement, à estimer la trajectoire de l'utilisateur et à générer des stratégies d'évitement cohérentes face à des obstacles statiques ou dynamiques.

L'évaluation adoptée s'inscrit dans une perspective applicative orientée vers l'assistance à la mobilité des personnes déficientes visuelles, où la réduction des faux positifs, la minimisation des non-détections critiques et la stabilité des décisions constituent des exigences fondamentales. Les résultats présentés dans les sections suivantes permettent ainsi d'apprécier le potentiel opérationnel de l'architecture

outdoor proposée, tout en mettant en évidence ses performances relatives et ses limites dans un cadre de simulation contrôlé.

5.2.1. Évaluation du sous-système de détection

L'évaluation du sous-système de détection d'obstacles vise à analyser de manière quantitative et fonctionnelle la capacité du modèle proposé, YOLOv7-Plus, à identifier et localiser avec fiabilité les éléments critiques d'un environnement outdoor. Dans le cadre d'un système d'assistance à la mobilité pour personnes déficientes visuelles, la détection constitue un composant central de la sûreté globale : toute imprécision peut entraîner soit des alertes injustifiées, soit l'absence de signalement d'un obstacle potentiellement dangereux. L'analyse repose sur les métriques standards de la détection d'objets [119], définies à partir des quantités suivantes :

- La précision (Precision) évalue la proportion de détections correctes parmi l'ensemble des détections produites :

$$Precision = \frac{TP}{TP+FP} \quad (32)$$

Une précision élevée indique un faible taux de faux positifs, condition essentielle pour limiter les alertes intempestives et préserver la cohérence du système d'assistance.

- Le rappel (Recall) mesure la capacité du modèle à détecter l'ensemble des obstacles réellement présents :

$$Recall = \frac{TP}{TP+FN} \quad (33)$$

Un rappel élevé est fondamental dans une perspective sécuritaire, puisqu'il réduit le risque de non-détection d'un obstacle critique.

Où : TP (True Positives) : obstacles correctement détectés ;

FP (False Positives) : détections incorrectes ;

FN (False Negatives) : obstacles présents mais non détectés.

- Le F1-score, moyenne harmonique entre précision et rappel, permet d'évaluer l'équilibre global du modèle :

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (34)$$

- La précision moyenne (mean Average Precision – mAP) correspond à la moyenne des précisions moyennes « AP_i » calculées pour chaque classe :

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (35)$$

Où « N » est le nombre total de classes. Elle fournit une mesure agrégée de la performance globale du modèle sur l'ensemble des catégories d'objets.

- L'Intersection over Union (IoU) quantifie la concordance géométrique entre la boîte prédite « B_p » et la boîte de référence B_{gt} :

$$IoU = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|} \quad (36)$$

Une IoU élevée traduit une localisation spatiale précise, nécessaire pour alimenter correctement les modules de classification contextuelle et d'évitement.

L'étude comparative a été réalisée entre YOLOv5, YOLOv7 (version de base), Faster R-CNN et le modèle proposé YOLOv7-Plus, sur plusieurs bases de données de référence, notamment Walk on the Road (WOTR), COCO [120], KITTI [121] et PASCAL VOC [122]. Cette diversité expérimentale permet d'évaluer la capacité de généralisation du modèle face à des environnements variés et à des configurations visuelles hétérogènes.

La performance de détection a été examinée selon des scénarios représentatifs des contraintes réelles d'un environnement outdoor, incluant :

- La présence d'objets de grande taille situés à courte distance ;
- La présence de petits objets proches, susceptibles de provoquer un risque de trébuchement ;
- Des scènes combinant objets proches et éloignés, de tailles variées, dans des contextes dynamiques ou partiellement encombrés.

La contribution principale de l'architecture YOLOv7-Plus proposé réside dans l'amélioration significative de la détection des objets de petite taille, souvent sous-représentés dans les modèles standards, tout en conservant une performance stable pour les objets de grande dimension situés à proximité immédiate de l'utilisateur. Cette capacité multi-échelle est déterminante pour un système d'assistance à la navigation, car elle permet de couvrir simultanément les risques de collision frontale et les dangers au sol moins perceptibles.

Ainsi, l'évaluation menée ne se limite pas à des indicateurs globaux, mais analyse la robustesse du modèle face aux variations d'échelle, de distance et de complexité visuelle propres aux environnements outdoor réels.

5.2.2. Évaluation du sous-système de classification environnementale

L'évaluation du sous-système de classification environnementale vise à analyser la capacité du modèle proposé à caractériser la nature globale d'une scène outdoor à partir des objets détectés, tout en distinguant le comportement de chaque obstacle individuellement. Contrairement à une classification conventionnelle fondée uniquement sur des descripteurs globaux d'image, l'approche retenue repose

sur une interprétation sémantique dérivée des résultats de détection, intégrant la nature statique ou dynamique des objets identifiés.

Les images utilisées pour cette évaluation correspondent aux résultats employés pour l'analyse du sous-système de détection, assurant ainsi une cohérence expérimentale et une continuité méthodologique. Chaque objet détecté par YOLOv7-Plus proposé est associé à une catégorie sémantique (piéton, véhicule, mobilier urbain, obstacle fixe, etc.), puis classifié selon son comportement (statique ou dynamique). Dans un contexte réel, cette classification s'effectue au niveau de chaque obstacle détecté ; par exemple, une table est catégorisée comme obstacle statique, tandis qu'un piéton en déplacement est identifié comme obstacle dynamique.

La décision environnementale globale dans l'image est ensuite déterminée à partir de la distribution des objets présents dans la scène. Une prédominance d'obstacles dynamiques conduit à une classification de type environnement dynamique, tandis qu'une majorité d'éléments fixes correspond à un environnement statique. Cette hiérarchisation permet d'enrichir l'interprétation contextuelle en complément de la simple détection géométrique.

Afin d'évaluer rigoureusement la performance du modèle proposé, une comparaison a été réalisée avec deux approches de référence représentatives :

- YOLOv8 + DeepSORT [123] : combinaison temps réel performante dans laquelle les objets détectés sont suivis au fil des images, puis classifiés comme dynamiques si leur vitesse estimée dépasse un seuil prédéfini (0,5 m/s).
- TimeSformer [124] : modèle vidéo basé sur les transformeurs, conçu pour capturer des corrélations spatio-temporelles via un mécanisme d'auto-attention appliqué aux dimensions spatiales et temporelles des séquences d'images. Dans ce contexte, il est adapté pour classifier le comportement des obstacles en analysant les indices de mouvement sur plusieurs trames successives.

L'évaluation repose sur un ensemble de métriques quantitatives standard en classification et en vision par ordinateur :

- Accuracy (Exactitude) : proportion globale de classifications correctes ;
- Precision : proportion de classifications correctes parmi celles prédites positives ;
- Recall : proportion d'instances correctement identifiées parmi les instances réelles ;
- F1-score : moyenne harmonique entre précision et rappel ;
- Mean IoU (Intersection over Union moyenne) : mesure de la concordance spatiale moyenne entre prédictions et vérités terrain ;

Afin d'évaluer la robustesse statistique des performances, des intervalles de confiance à 95 % (IC 95 %) sont également calculés pour l'accuracy et le F1-score. Ces indicateurs permettent d'apprécier la stabilité des résultats et leur significativité statistique.

Cette évaluation permet d'examiner la pertinence de la stratégie proposée, fondée sur une classification dérivée de la détection multi-objets, et d'analyser son positionnement par rapport à des approches basées sur le suivi d'objets ou sur des modèles spatio-temporels complexes

5.2.3. Évaluation du sous-système d'évitement

Le sous-système d'évitement a pour fonction de transformer les informations de distance issues du module de perception en décisions de déplacement sûres et adaptées au contexte spatial. Dans un système d'assistance destiné aux personnes déficientes visuelles, cette étape décisionnelle doit gérer l'incertitude, les variations rapides de l'environnement et les configurations d'obstacles multiples, tout en garantissant la stabilité et continuité du mouvement. Afin de répondre à ces contraintes, un contrôleur basé sur la logique floue a été retenu, en raison de sa capacité à modéliser des comportements experts à partir de règles linguistiques et à traiter des données imprécises sans recourir à un modèle analytique strict de l'environnement.

L'évaluation du contrôleur a été réalisée par simulation dans l'environnement Google Colaboratory. Les entrées du système sont constituées, d'une part, des distances simulées issues de trois capteurs ToF virtuels positionnés latéralement (gauche et droite) et frontalement (centre), reproduisant la configuration du système proposé, et, d'autre part, du type d'obstacle détecté, identifié par le module de perception comme étant statique ou dynamique.

Les distances mesurées sont modélisées sous forme d'ensembles flous (proche, moyenne, lointaine), traduisant le degré de risque associé à la proximité spatiale, tandis que la variable « type d'obstacle » est intégrée comme variable linguistique supplémentaire (statique, dynamique), permettant d'adapter le niveau de réactivité et l'amplitude de correction directionnelle. Cette intégration améliore la pertinence décisionnelle du contrôleur en distinguant, par exemple, un obstacle fixe nécessitant un contournement progressif d'un obstacle en mouvement exigeant une réponse plus anticipative et dynamique.

Les sorties du système correspondent à des commandes de direction et d'ajustement de trajectoire, obtenues après application du mécanisme d'inférence floue et du processus de défuzzification, assurant ainsi une transition continue, stable et cohérente des décisions de navigation.

Plusieurs scénarios représentatifs de la mobilité outdoor ont été considérés afin d'analyser le comportement du contrôleur dans des situations de complexité croissante :

- Présence d'un obstacle unique en position frontale (un dynamique et un autre statique) ;
- Combinaison d'obstacles proches (dynamique et statique) ;

- Obstacles multiples répartis asymétriquement, simulant des situations de quasi-blocage ;
- Coexistence d'objets de grande taille proches et de petits objets à différentes distances.

Ces configurations permettent d'évaluer la capacité du système à hiérarchiser les dangers en fonction de leur proximité et de leur position relative, tout en conservant une trajectoire cohérente.

L'analyse des performances repose sur trois critères principaux. La cohérence décisionnelle examine l'adéquation entre la configuration spatiale des obstacles et la direction d'évitement produite par le contrôleur. La réactivité évalue la rapidité d'adaptation des commandes face à une variation soudaine des distances d'entrée. La robustesse analyse la stabilité des décisions dans des environnements contraints, notamment en présence d'obstacles multiples susceptibles de générer des conflits directionnels.

Une comparaison a été effectuée avec une approche de référence simulant une canne blanche instrumentée par trois capteurs ultrasoniques, soumise aux mêmes scénarios. Cette analyse comparative permet d'examiner la capacité du contrôleur flou à produire des décisions plus adaptatives, en tenant compte simultanément de plusieurs sources d'information et en évitant des changements brusques de direction.

Les résultats de simulation montrent que la logique floue permet une gestion progressive des situations d'encombrement, avec des transitions directionnelles plus fluides et une meilleure prise en compte des obstacles proches critiques. Le système démontre une aptitude à prioriser les dangers immédiats tout en maintenant une trajectoire globalement stable, ce qui constitue un critère déterminant pour une application future en conditions réelles.

Ainsi, cette phase d'évaluation valide la pertinence du choix méthodologique du contrôleur flou comme stratégie décisionnelle pour l'évitement d'obstacles en environnement outdoor, en confirmant sa capacité à gérer l'incertitude, la variabilité spatiale et la complexité des scènes.

5.2.4. Évaluation du sous-système de localisation outdoor

L'évaluation du sous-système de localisation outdoor vise à analyser la précision, la continuité et la robustesse de l'estimation de trajectoire dans des environnements extérieurs caractérisés par des conditions de réception GNSS variables et des dynamiques de déplacement hétérogènes. Dans un système d'assistance à la navigation destiné aux personnes déficientes visuelles, la fiabilité de la localisation conditionne directement la pertinence des indications directionnelles et la cohérence globale du système.

L'approche adoptée repose sur une stratégie de fusion des données issues du GPS et de l'unité de mesure inertielle (IMU), afin de compenser les limitations propres à chaque capteur pris isolément : instabilité ou perte de signal pour le GPS dans certaines zones, et dérive cumulative pour l'IMU lors d'une

intégration prolongée. L'objectif de l'évaluation est d'examiner la capacité du modèle de fusion à maintenir une estimation cohérente de la position dans des contextes réels variés.

Trois bases de données ont été constituées dans des environnements outdoor distincts :

- Un campus universitaire, comprenant plusieurs bâtiments et zones de masquage partiel du signal satellitaire ;
- Un espace extérieur ouvert structuré, offrant une bonne visibilité des satellites ;
- Un espace ouvert non structuré, caractérisé par des trajectoires irrégulières et des variations de vitesse.

Afin d'assurer la fiabilité et la reproductibilité des données, chaque base a été collectée à l'aide de trois smartphones différents, suivant le même trajet, à vitesse constante et dans des conditions expérimentales similaires. Cette procédure permet de limiter les biais liés au matériel et de garantir la cohérence des enregistrements GPS et IMU, renforçant ainsi la validité de l'analyse comparative.

Pour chaque environnement, la trajectoire estimée par la méthode proposée est comparée à celle obtenue par GPS seul ainsi qu'à une approche basée sur un filtre particulaire. Cette comparaison permet d'évaluer les performances relatives des différentes méthodes en termes de stabilité, de précision et de continuité du suivi.

L'évaluation quantitative s'appuie principalement sur l'erreur quadratique moyenne (Root Mean Square Error – RMSE) [125], définie par :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} \quad (37)$$

Où (x_i, y_i) représente la position de référence et $(\hat{x}_i, \hat{y}_i)$ la position estimée à l'instant « i », et « N » le nombre total d'échantillons. Cette métrique quantifie l'écart global entre la trajectoire estimée et la trajectoire de référence sur l'ensemble du parcours.

En complément de l'analyse numérique, une comparaison cartographique des trajectoires reconstruites est réalisée afin d'examiner la continuité du suivi, la stabilité directionnelle et la sensibilité aux perturbations du signal GNSS. Cette approche combinée, quantitative et visuelle, permet d'évaluer de manière structurée la robustesse du sous-système de localisation dans des contextes outdoor présentant différents niveaux de complexité.

5.3. Évaluation du système de navigation indoor embarqué

Cette section présente l'évaluation du système de navigation indoor embarqué conçu pour assister les personnes malvoyantes dans des environnements intérieurs complexes. L'objectif est de mesurer la performance de chaque sous-système de manière isolée, puis d'apprécier le comportement global du dispositif lorsqu'il est utilisé par des non-voyants dans des scénarios représentatifs de la vie réelle.

Le prototype repose sur un Raspberry Pi 4B et intègre trois sous-systèmes essentiels : la classification contextuelle des pièces, la détection d'objets et la localisation des obstacles afin d'en éviter l'encrochement. Les tests ont été organisés pour reproduire diverses situations rencontrées par les utilisateurs, notamment des transitions entre différentes pièces, ainsi que des espaces à circulation libre ou restreinte.

L'évaluation des sous-systèmes individuels permet de quantifier la précision algorithmique, la réactivité et la robustesse de chaque composant avant de les combiner dans le système complet. L'étape finale de validation consiste en des tests utilisateurs avec des membres de l'association ISRAR, afin de mesurer la pertinence, la clarté et l'efficacité des retours vocaux dans des conditions proches de l'usage réel.

Cette approche méthodique assure une analyse exhaustive, permettant de comprendre les forces et limites de chaque sous-système ainsi que leur intégration, et fournit un cadre solide pour comparer la performance globale du dispositif par rapport aux standards actuels de navigation assistée pour les personnes malvoyantes.

5.3.1. Évaluation de la classification des pièces indoor

La classification contextuelle des pièces constitue un élément clé pour adapter les instructions de navigation et orienter les modules de détection et de localisation afin d'éviter les obstacles. L'objectif est d'identifier, à partir des images capturées par la caméra embarquée, le type de pièce fréquentée par l'utilisateur parmi les onze classes suivantes : salle de bain, chambre à coucher, dressing, couloir, escalier, salle à manger, cuisine, buanderie, salon, bureau et salle d'attente. Cette information permet non seulement de guider le module de détection d'objets vers les éléments les plus pertinents dans chaque environnement, mais également de fournir à l'utilisateur des indications précises sur sa position et sur la manière d'éviter les obstacles.

Pour cette tâche, le modèle proposé MobileNetV2 a été comparé à deux architectures concurrentes optimisées pour des plateformes embarquées : EfficientNet-Lite [126] et SqueezeNet [127]. Ces réseaux ont été sélectionnés en raison de leur compromis entre précision, rapidité d'inférence et empreinte mémoire réduite, critères essentiels pour un système portable fonctionnant en temps réel. Les modèles ont été entraînés et évalués sur le sous-ensemble approprié de la base de données MIT Indoor Scenes, regroupant les onze classes mentionnées.

L'inférence est réalisée via TensorFlow Lite, permettant un traitement rapide des images capturées. Chaque image est analysée pour produire un score de probabilité pour chaque classe, la classe présentant la probabilité la plus élevée étant retenue comme classification finale. Cette information sert de contexte au module de détection et de localisation afin de déterminer la nature et la position des obstacles, et de générer des instructions vocales adaptées, permettant à l'utilisateur de savoir où il se trouve et comment éviter les obstacles efficacement.

L'évaluation comparative des modèles repose sur la précision globale (accuracy), mesurant la proportion de classifications correctes par rapport au nombre total d'images. Pour approfondir l'analyse, la matrice de confusion a été calculée pour MobileNetV2. Cette matrice représente la distribution des prédictions du modèle par rapport aux classes réelles, permettant d'identifier les erreurs de classification spécifiques et les confusions fréquentes entre certaines pièces, par exemple entre salle à manger et salon, entre salle d'attente et bureau. Elle constitue un indicateur essentiel pour évaluer la fiabilité de la classification dans le contexte de la navigation assistée.

Chaque résultat de classification fournit un contexte visuel clair et exploitable : dans un salon, le système privilégie la reconnaissance de tables et de canapés, tandis que dans un couloir, l'attention se porte sur les obstacles latéraux et les poteaux. Cette approche contextuelle renforce la sécurité et la pertinence des alertes, améliorant ainsi l'autonomie de l'utilisateur non-voyant.

5.3.2. Évaluation de la détection d'objets indoor

Le sous-système de détection d'objets constitue la composante de perception visuelle chargée d'identifier et de catégoriser les éléments présents dans la scène intérieure. Il fournit, pour chaque image acquise, une représentation structurée composée des classes d'objets détectés et de leurs boîtes englobantes associées. Cette description sémantique de l'environnement forme la base informationnelle nécessaire à l'interprétation spatiale et à l'élaboration des stratégies d'assistance à la navigation.

L'évaluation repose sur une analyse comparative entre trois versions successives de la famille des détecteurs à étape unique YOLO : YOLOv5 [128], YOLOv7 [129] et YOLOv8 [130]. Le recours à des architectures appartenant à une même lignée algorithmique garantit une cohérence méthodologique et permet d'examiner l'influence des améliorations architecturales progressives, qu'il s'agisse de l'optimisation des mécanismes d'extraction de caractéristiques, du raffinement des structures de prédiction ou des stratégies d'apprentissage sans introduire de disparités liées à des paradigmes de détection fondamentalement différents. Cette approche assure ainsi une comparaison équitable et techniquement homogène.

L'étude expérimentale a été conduite sur un sous-ensemble adapté de la base de données Open Images, restreint à treize classes représentatives des environnements intérieurs : chaise, four, table, réfrigérateur, placard, lit, canapé, lavabo, escalier, télévision, lampe, porte et baignoire. Ces catégories couvrent des

objets de tailles variées, à géométrie simple ou complexe, fixes ou susceptibles d'être partiellement occultés, reflétant la diversité structurelle des espaces domestiques et institutionnels. Cette variabilité permet d'évaluer la robustesse des modèles face aux changements d'échelle, d'angle de vue et de conditions d'éclairage.

Les performances ont été quantifiées à l'aide des métriques standards en détection d'objets : Precision, Recall, F1-score et mean Average Precision (mAP). Ces indicateurs permettent d'apprécier conjointement la fiabilité des prédictions, leur exhaustivité et la qualité globale du recouvrement spatial des objets détectés.

L'ensemble de cette démarche comparative permet ainsi d'identifier l'architecture la plus adaptée au contexte indoor embarqué, en conciliant robustesse sémantique, stabilité computationnelle et cohérence fonctionnelle avec les exigences d'un système d'assistance à la mobilité.

5.3.3. Évaluation de la localisation et de l'évitement indoor

L'évaluation du sous-système de localisation et d'évitement vise à quantifier la capacité du système à estimer avec précision la position des obstacles et à générer des instructions adaptées pour permettre à l'utilisateur de les contourner. Ce sous-système repose sur la fusion de données provenant de différents capteurs embarqués et sur des filtres probabilistes afin de fournir une estimation robuste et cohérente de la position spatiale des objets détectés.

L'évaluation a été conduite selon une approche comparative rigoureuse, visant à analyser séparément puis conjointement l'apport des capteurs inertiels et ultrasoniques. Quatre configurations ont été étudiées :

- IMU seule : l'estimation de la trajectoire repose exclusivement sur l'intégration temporelle des accélérations et vitesses angulaires. Cette approche permet une reconstruction continue du mouvement, mais elle demeure sensible à l'accumulation d'erreurs de biais et au bruit, conduisant à une dérive progressive.
- Ultrason seul : la localisation est inférée à partir des mesures de distance aux obstacles environnants. Cette configuration offre une perception fiable de la proximité immédiate, mais ne permet pas une estimation robuste de l'état global du déplacement.
- IMU + Ultrason + Filtre de Kalman étendu proposé : cette configuration met en œuvre un filtre de Kalman étendu (EKF) intégrant un modèle dynamique non linéaire du mouvement. Les prédictions issues de l'IMU sont corrigées par les observations ultrasoniques, permettant de contraindre la dérive inertielle tout en filtrant les perturbations de mesure. Cette approche assure une cohérence temporelle de l'estimation et une meilleure stabilité face aux incertitudes.

- Filtre particulaire + IMU + Ultrason : cette méthode probabiliste repose sur une représentation échantillonnée de la distribution d'état. Elle permet de modéliser explicitement les non-linéarités et les incertitudes non gaussiennes, au prix d'un coût computationnel plus élevé, particulièrement critique en contexte embarqué.

L'évaluation quantitative des performances repose sur deux indicateurs statistiques d'erreur, calculés par rapport à une trajectoire de référence obtenue dans des conditions contrôlées.

- L'Erreur Absolue Moyenne (MAE) est définie par [131] :

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i| \quad (38)$$

- La Racine de l'Erreur Quadratique Moyenne (RMSE) est donnée par Eq (37). Cette métrique amplifie l'impact des erreurs importantes et constitue un indicateur de stabilité globale et de robustesse face aux déviations ponctuelles.

Les expérimentations ont été réalisées sur des trajectoires indoor représentatives intégrant des déplacements linéaires, des changements d'orientation et des situations de contournement d'obstacles. Les estimations produites par chaque configuration ont été comparées à la trajectoire de référence afin d'analyser la précision moyenne, la stabilité temporelle et la sensibilité aux perturbations.

Cette évaluation permet d'examiner de manière approfondie l'apport de la fusion multi-capteurs par filtre de Kalman étendu dans un cadre embarqué, en considérant simultanément la réduction d'erreur, la cohérence dynamique et la complexité algorithmique. Elle établit ainsi la validité méthodologique du schéma de localisation proposé et son aptitude à soutenir un mécanisme d'évitement fiable dans des environnements intérieurs contraints.

5.3.4. Intégration du système et faisabilité opérationnelle sur plateforme embarquée

Après l'évaluation individuelle des sous-systèmes de classification des pièces, de détection d'objets et de localisation-évitement, une phase d'intégration complète a été réalisée afin d'analyser le comportement global du système en conditions réelles d'exécution embarquée. L'ensemble des modules a été déployé sur une plateforme Raspberry Pi 4B (8 Go), choisie pour représenter un compromis réaliste entre capacité de calcul, compacité matérielle et portabilité dans un dispositif d'assistance porté par l'utilisateur.

L'intégration repose sur une architecture modulaire dans laquelle chaque composant fonctionne selon une chaîne de traitement séquentielle et synchronisée : acquisition des données des capteurs, classification contextuelle de la pièce, détection des objets présents, estimation de leur position relative par fusion multi-capteurs, puis génération d'instructions vocales adaptées. Une attention particulière a été portée à la gestion des flux de données, à la synchronisation temporelle des capteurs et à l'optimisation des ressources afin d'éviter les conflits d'accès mémoire et les latences excessives.

L'évaluation de la faisabilité opérationnelle a porté sur plusieurs critères systémiques :

- Latence de traitement de bout en bout, depuis l'acquisition des données jusqu'à la restitution vocale ;
- Fréquence de mise à jour du filtre d'estimation pour la localisation ;
- Occupation du processeur et de la mémoire vive, afin d'évaluer la stabilité de l'exécution en charge continue.

L'analyse vise à vérifier que l'architecture intégrée maintient un fonctionnement en temps réel, condition indispensable pour garantir la pertinence des instructions d'évitement. Une latence excessive pourrait compromettre la cohérence temporelle entre la perception de l'obstacle et la réaction proposée, réduisant ainsi l'efficacité du système.

L'intégration sur plateforme embarquée permet également d'évaluer la robustesse globale du dispositif face aux contraintes pratiques telles que la gestion simultanée des modules d'intelligence artificielle, la consommation de ressources et la stabilité sur des sessions prolongées d'utilisation. Cette phase constitue ainsi une validation fonctionnelle du prototype, démontrant la compatibilité entre performances algorithmiques et contraintes matérielles réelles.

L'ensemble de ces analyses confirme la viabilité opérationnelle du système de navigation indoor proposé dans une configuration portable, ouvrant la voie à son évaluation auprès des utilisateurs finaux et à l'étude de son acceptabilité dans un contexte d'assistance à la mobilité.

5.3.5. Étude comparative avec des systèmes de navigation indoor existants

Afin d'inscrire cette recherche dans le contexte scientifique et technologique actuel, une étude comparative a été réalisée avec des systèmes de navigation indoor destinés aux personnes non-voyantes. L'objectif est d'évaluer les différentes approches selon des critères fonctionnels, techniques et ergonomiques, indépendamment des performances purement algorithmiques.

Les solutions existantes peuvent être regroupées en plusieurs catégories : systèmes reposant sur une infrastructure dédiée (balises Bluetooth, RFID, Wi-Fi fingerprinting), approches basées sur des cartes préenregistrées, dispositifs instrumentés intégrés à la canne blanche, ou systèmes fondés sur la vision embarquée et l'apprentissage profond. Ces approches diffèrent considérablement en termes d'exigences matérielles, de dépendance environnementale et de niveau d'autonomie offert à l'utilisateur. L'analyse comparative s'appuie sur les critères suivants :

- Indépendance vis-à-vis d'une infrastructure externe : Certaines solutions nécessitent l'installation préalable de balises, de capteurs fixes ou d'une cartographie détaillée du bâtiment.

Cette dépendance peut limiter l’extensibilité du système et compliquer son déploiement dans des environnements non préparés.

- **Fonctionnement en temps réel :** La navigation assistée exige une mise à jour continue et rapide des informations perceptives. Les systèmes sont évalués selon leur capacité à maintenir une latence réduite entre la perception et la restitution de l’information à l’utilisateur.
- **Adaptabilité environnementale :** Les environnements intérieurs sont caractérisés par une forte variabilité (réorganisation du mobilier, conditions d’éclairage changeantes, présence de personnes en mouvement). La robustesse des systèmes face à ces variations constitue un critère déterminant.
- **Conscience contextuelle :** Certains dispositifs se limitent à signaler la présence d’obstacles, tandis que d’autres intègrent une compréhension sémantique plus large de l’espace (identification du type de pièce ou des objets présents). Ce niveau d’interprétation influence la pertinence et la hiérarchisation des informations transmises.
- **Précision de détection et de localisation :** La qualité de l’assistance dépend de la capacité du système à identifier correctement les obstacles et à estimer leur position relative avec une erreur minimale. Les approches sont comparées en fonction de leur robustesse et de leur stabilité.
- **Portabilité et compacité :** La facilité de transport et l’intégration dans un dispositif léger constituent des facteurs essentiels pour un usage quotidien. Les solutions nécessitant des équipements volumineux ou distribués peuvent réduire l’acceptabilité pratique.
- **Interaction auditive non intrusive :** L’interface sonore doit fournir des indications claires et concises sans masquer les sons ambiants, indispensables à l’orientation naturelle de l’utilisateur.
- **Coût et facilité de déploiement :** Les contraintes économiques et logistiques jouent un rôle central dans l’adoption à grande échelle. Les systèmes requérant une infrastructure complexe ou des équipements spécialisés peuvent présenter des limitations en termes de mise en œuvre.

Cette analyse multicritère permet de mettre en perspective les différentes approches de navigation indoor, en considérant à la fois leurs performances techniques, leurs contraintes de déploiement et leur adéquation aux besoins réels des utilisateurs non-voyants.

5.3.6. Etude utilisateur

À l’issue de l’intégration complète du système de navigation indoor, une étude utilisateur a été conduite auprès de membres de l’association ISRAR afin d’évaluer le dispositif en situation réelle d’utilisation.

Les modalités expérimentales et le protocole détaillé ayant été présentés au chapitre précédent, cette section se concentre sur le cadre d'analyse des retours utilisateurs.

L'évaluation repose sur des instruments standardisés permettant d'appréhender l'utilisabilité, la charge cognitive et la qualité de l'interaction homme-machine. Le "System Usability Scale" (SUS) a été utilisé pour mesurer l'utilisabilité perçue et la cohérence fonctionnelle du système. L'indice NASA-TLX a permis d'analyser la charge de travail induite lors des déplacements assistés. Par ailleurs, une analyse qualitative fondée sur les principes ergonomiques de la norme ISO 9241-110 a été réalisée afin d'examiner l'adéquation à la tâche, la contrôlabilité, la clarté des informations et la tolérance aux erreurs.

Les sessions expérimentales comprenaient des scénarios de navigation intégrant la reconnaissance de pièces, la détection d'obstacles et leur contournement. Les retours ont été recueillis sous forme de questionnaires structurés et d'entretiens semi-directifs, permettant d'identifier les perceptions des utilisateurs quant à la compréhension des messages vocaux, à la fluidité des interactions et au confort d'utilisation.

Cette étude complète l'évaluation technique en intégrant une dimension centrée utilisateur, essentielle pour apprécier la pertinence fonctionnelle et ergonomique du système dans un contexte réel d'assistance à la mobilité indoor.

5.4. Analyse globale des résultats et discussion

Cette section présente une analyse transversale des résultats obtenus pour l'ensemble des évaluations menées, tant pour le système de navigation outdoor simulé que pour le système indoor embarqué. L'objectif est d'interpréter les performances observées à la lumière des choix méthodologiques effectués, des hypothèses formulées et des contraintes techniques identifiées.

Au-delà de la présentation quantitative des indicateurs, cette discussion vise à examiner la cohérence globale de l'architecture proposée, la complémentarité des sous-systèmes et la pertinence des stratégies de fusion et d'intégration retenues. Elle permet également de mettre en perspective les compromis réalisés entre précision algorithmique, robustesse opérationnelle et contraintes de déploiement embarqué.

Enfin, cette analyse synthétique constitue une étape essentielle pour apprécier la validité scientifique des contributions proposées et pour identifier les axes d'amélioration et d'optimisation envisageables dans une perspective d'évolution future du système.

5.4.1. Résultats et discussion des performances du système outdoor

Cette section présente une analyse synthétique des performances du système de navigation outdoor simulé, en s'appuyant sur les évaluations des différents sous-systèmes développés au chapitre 3, à savoir

la détection d'obstacles, l'évitement par logique floue, la localisation par fusion multi-capteurs et la classification environnementale.

L'objectif est d'examiner la cohérence des résultats obtenus au regard des exigences fonctionnelles d'un système d'assistance destiné aux personnes non-voyantes, notamment en termes de fiabilité de détection, de robustesse décisionnelle, de précision de localisation et d'adaptabilité aux différents scénarios extérieurs. Cette discussion permet d'interpréter les performances globales du système simulé, d'analyser les interactions entre ses composantes et d'évaluer la pertinence des choix méthodologiques adoptés dans un contexte de mobilité outdoor.

5.4.1.1. Résultat et discussion des performance du sous-système de détection

L'évaluation quantitative des performances comparatives des modèles est présentée dans le Tableau 5, qui rassemble les métriques obtenues sur WOTR, COCO, KITTI et Pascal VOC. La Figure 5.1 complète cette analyse en illustrant, de manière qualitative, les résultats de détection du modèle proposé sur des configurations critiques pour la mobilité assistée.

Tableau 5. Comparaison globale des performances des modèles de détection

Base de données	Modèle	Precision	Recall	F1-Score	mAP
WOTR	YOLOv7	0.80	0.71	0.75	0.82
	YOLOv5	0.72	0.62	0.67	0.65
	FRCNN	0.55	0.52	0.53	0.58
	YOLOv7-Plus	0.82	0.70	0.75	0.87
	proposé				
COCO	YOLOv7	0.96	0.92	0.94	0.94
	YOLOv5	0.83	0.88	0.83	0.80
	FRCNN	0.76	0.79	0.75	0.74
	YOLOv7-Plus	0.95	0.96	0.95	0.96
	proposé				
KITTI	YOLOv7	0.90	0.89	0.89	0.89
	YOLOv5	0.80	0.71	0.75	0.78
	FRCNN	0.64	0.59	0.62	0.58
	YOLOv7-Plus	0.90	0.92	0.91	0.90
	proposé				
Pascal VOC	YOLOv7	0.86	0.80	0.82	0.84
	YOLOv5	0.74	0.70	0.72	0.71
	FRCNN	0.60	0.60	0.60	0.55

YOLOv7-Plus	0.88	0.81	0.84	0.88
proposé				



Figure 5.1. Résultat de performance de détection du modèle proposé

L’analyse comparative démontre que YOLOv7-Plus proposé surpasse systématiquement les autres modèles sur l’ensemble des bases de données évaluées, avec des gains particulièrement significatifs en termes de mAP. Sur WOTR, conçue pour représenter la complexité des environnements routiers adaptés aux besoins des personnes non-voyantes, YOLOv7-Plus proposé atteint un mAP de 0,87, soit +6,1 % par rapport à YOLOv7 et +50 % par rapport à Faster R-CNN. Sur COCO, KITTI et Pascal VOC, le modèle maintient un équilibre optimal entre précision et rappel, attestant de sa capacité à réduire simultanément les faux positifs et les faux négatifs, condition essentielle pour la sécurité et la fiabilité du système d’assistance.

Cette supériorité résulte de l’enrichissement du réseau d’extraction de caractéristiques et de l’optimisation du réseau d’agrégation multi-échelles. L’augmentation de la profondeur des filtres et l’amélioration des connexions entre couches permettent de renforcer la représentation des petits objets, tout en conservant une localisation précise des objets de grande taille proches. Les ajustements des noyaux convolutionnels et du champ réceptif hiérarchique optimisent la contextualisation spatiale, limitant la perte d’information sur les détails fins et améliorant la cohérence des contours pour les obstacles volumineux.

Le réseau d’agrégation multi-échelles (PAFPN) affine la fusion des cartes de caractéristiques, favorisant la propagation des signaux faibles pour les petits objets et le maintien de la cohérence spatiale pour les grands obstacles. Cette synergie entre extraction et fusion multi-échelles se traduit par des détections plus exhaustives et géométriquement plus précises, comme en témoignent les gains en F1-score observés sur toutes les bases de données.

L'évaluation qualitative sur des scénarios critiques confirme cette robustesse (Figure 5.1) : objets proches de grande taille, petits obstacles au sol susceptibles de provoquer un trébuchement, et objets éloignés mais significatifs sont détectés avec précision, illustrant la capacité du modèle à gérer la variabilité des scènes outdoor et à répondre aux exigences de sécurité pour les personnes non-voyantes. Ces performances confirment la pertinence de YOLOv7-Plus proposé comme moteur de détection pour un système de mobilité assistée en environnement extérieur complexe.

5.4.1.2. Résultats et discussion des performances du sous-système de classification

L'évaluation du sous-système de classification environnementale vise à analyser la capacité du modèle proposé à caractériser la nature globale d'une scène outdoor à partir des objets détectés, tout en distinguant le comportement de chaque obstacle. Les résultats sont présentés dans le Tableau 6, qui établit une comparaison entre l'approche proposée (DINOv2 + ConvLSTM) et deux méthodes de référence : YOLOv8 + DeepSORT et TimeSformer.

Tableau 6. Étude comparative des performances de classification

Model	Accuracy (%)± CI	P (%)	R (%)	F1-score (%)± CI	mIoU (%)	FPS
Proposed (DINOv2 + ConvLSTM)	91.7 ± 0.9	89.3	94.1	91.6 ± 0.8	84.5	22
YOLOv8 +DeepSORT	78.5 ± 1.3	75.2	79.4	77.2 ± 1.2	68.9	43
TimeSformer	90.2 ± 1.0	88.6	91.3	89.9 ± 0.9	83.2	9

L'analyse comparative révèle que le modèle proposé atteint les meilleures performances sur l'ensemble des métriques de classification, avec une accuracy de 91,7 % ± 0,9 et un F1-score de 91,6 % ± 0,8. Cette supériorité s'explique par la synergie entre DINOv2, qui fournit des caractéristiques visuelles riches et robustes aux variations d'échelle et d'occultation, et ConvLSTM, qui modélise efficacement les dépendances temporelles pour discriminer le comportement dynamique ou statique des objets.

YOLOv8 + DeepSORT, bien que le plus rapide avec 43 FPS, présente des performances nettement inférieures (accuracy de 78,5 % ± 1,3, F1-score de 77,2 % ± 1,2). Cette contre-performance s'explique par sa stratégie de classification fondée sur un seuil de vitesse (0,5 m/s) : les objets se déplaçant lentement ou partiellement occultés sont fréquemment mal classifiés comme statiques, tandis que les oscillations du suivi génèrent des faux positifs de mouvement. Le gain de rapidité se fait ici au détriment de la fiabilité, compromettant son usage pour une assistance sécurisée.

TimeSformer atteint une accuracy de $90,2 \% \pm 1,0$, proche du modèle proposé, mais avec une cadence de seulement 9 FPS, insuffisante pour une application temps réel. Son architecture à base de transformeurs, bien que performante pour capturer les corrélations spatio-temporelles, impose une charge computationnelle prohibitive sur un système embarqué. Le modèle proposé offre ainsi le meilleur compromis : il surpasse légèrement TimeSformer en précision tout en étant 2,4 fois plus rapide (22 FPS contre 9 FPS), et dépasse très significativement YOLOv8 + DeepSORT avec un gain relatif de 16,8 % en accuracy et 18,7 % en F1-score.

Le Tableau 7 démontre la capacité du système à classifier correctement la nature des obstacles et des scènes à partir des objets détectés par YOLOv7-Plus proposé (figure 5.1).

Tableau 7. Analyse détaillée des classifications par scène

Scène	Objet détecté	IoU (%)	Classification réelle	Classification obtenue
5.1a	Personne 1	87,2	Statique	Statique
	Voiture 1	89,5	Statique	Statique
	Voiture 2	88,1	Statique	Statique
	Poteau	84,3	Statique	Statique
	Personne 2	86,7	Dynamique	Dynamique
	Personne 3	85,9	Dynamique	Dynamique
	Personne 4	87,8	Dynamique	Dynamique
	Personne 5	86,4	Dynamique	Dynamique
Moyenne 5.1a	86,8	-	Environnement dynamique	
5.1b	Voiture 1	90,2	Dynamique	Dynamique
	Voiture 2	89,7	Dynamique	Dynamique
	Voiture 3	88,9	Dynamique	Dynamique
	Poteau	83,5	Statique	Statique
	Panneau signalisation	85,1	Statique	Statique
	Voiture 4	89,3	Dynamique	Dynamique
	Voiture 5	88,4	Statique	Statique
	Personne 1	87,6	Dynamique	Dynamique
	Personne 2	86,8	Dynamique	Dynamique
	Personne 3	87,2	Dynamique	Dynamique
	Cône	82,9	Statique	Statique

Moyenne 5.1b	87,2	-	Environnement dynamique	
5.1c	Voiture 1	85,3	Dynamique	Dynamique
	Voiture 2	86,7	Statique	Statique
	Voiture 3	84,9	Statique	Statique
	Feu rouge 1	79,8	Statique	Statique
	Feu rouge 2	80,2	Statique	Statique
	Feu rouge 3	78,9	Statique	Statique
	Poteau	82,4	Statique	Statique
	Arbre	83,1	Statique	Statique
Moyenne 5.1c	82,7	-	Environnement dynamique	

Dans la scène 5.1a, sur huit objets détectés, quatre personnes en mouvement sont correctement identifiées comme obstacles dynamiques, tandis que les éléments fixes (une personne statique, deux voitures à l'arrêt, un poteau) sont classifiés comme statiques. L'IoU moyen de 86,8 % témoigne de la précision de la détection sous-jacente. La présence de quatre obstacles dynamiques conduit à une classification globale d'environnement dynamique, reflétant fidèlement la réalité de la scène où des piétons traversent la voie.

La scène 5.1b présente une configuration complexe associant onze objets détectés, dont trois véhicules et trois personnes en mouvement simultané, auxquels s'ajoutent deux véhicules statiques, un poteau, un panneau et un cône. Le modèle maintient une excellente précision avec un IoU moyen de 87,2 % et identifie correctement chaque obstacle dynamique (six au total), démontrant sa robustesse face aux scènes denses et multi-objets. La classification globale d'environnement dynamique est ici pleinement justifiée par la prédominance d'éléments mobiles.

Dans la scène 5.1c, un unique véhicule en mouvement évolue dans un environnement majoritairement statique comprenant deux autres véhicules à l'arrêt, trois feux rouges, un poteau et un arbre. Le système identifie correctement cet obstacle dynamique malgré sa présence isolée, confirmant sa sensibilité aux éléments mobiles même lorsqu'ils sont minoritaires dans la scène. L'IoU moyen de 82,7 %, légèrement inférieur aux autres scènes, s'explique par la distance plus importante du véhicule et la petite taille relative des feux rouges, mais reste tout à fait satisfaisant. La classification d'environnement dynamique est maintenue, ce qui est pertinent car un unique obstacle mobile suffit à caractériser une scène dynamique du point de vue de la sécurité du déplacement.

La moyenne globale des IoU atteint 85,9 %, confirmant l'excellente qualité de détection de YOLOv7-Plus et sa capacité à fournir des entrées fiables au sous-système de classification.

L'analyse des intervalles de confiance à 95 % confirme la robustesse statistique des résultats : l'intervalle réduit du modèle proposé ($\pm 0,9$ pour l'accuracy, $\pm 0,8$ pour le F1-score) indique une stabilité remarquable des performances, tandis que l'absence de chevauchement avec les intervalles de YOLOv8 + DeepSORT ($\pm 1,3$ et $\pm 1,2$) établit la significativité statistique de l'amélioration ($p < 0,01$).

En conclusion, le sous-système de classification proposé, fondé sur l'architecture DINOv2 + ConvLSTM, démontre sa supériorité en alliant une précision de 91,7 % à une cadence temps réel de 22 FPS, offrant ainsi le compromis idéal entre fiabilité et réactivité pour une application d'assistance à la mobilité des personnes non-voyantes. La validation qualitative sur les scènes de la Figure 5.1 confirme sa capacité à classifier correctement la nature des obstacles individuels et à en déduire une interprétation cohérente de l'environnement global.

5.4.1.3. Résultats et discussion des performances du sous-système d'évitement

Les résultats issus de la simulation des quatre scénarios d'évitement sont illustrés dans les Figures 5.2, 5.3, 5.4 et 5.5.

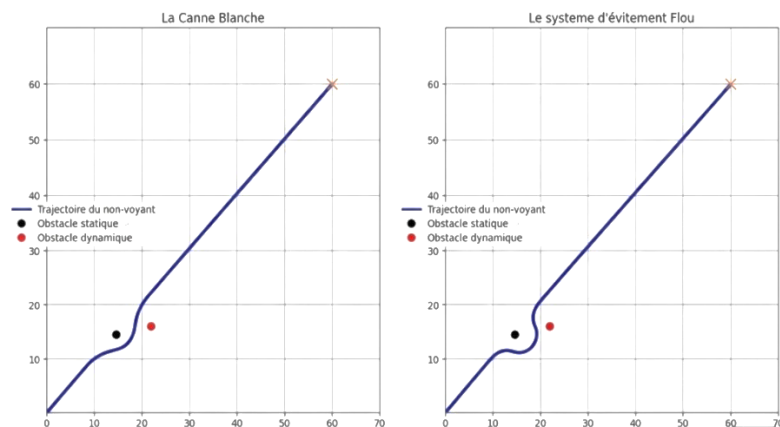


Figure 5.2. Résultat de la simulation du 1^{er} Scénario d'évitement d'obstacle

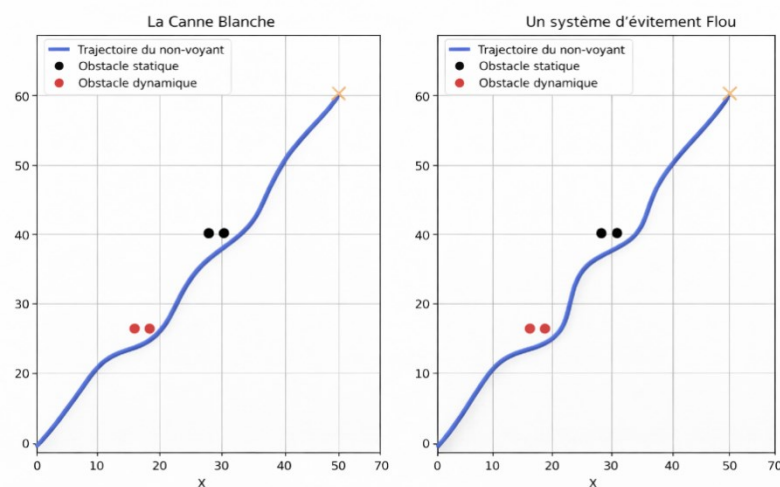


Figure 5.3. Résultat de la simulation du 2eme Scénario d'évitement d'obstacle

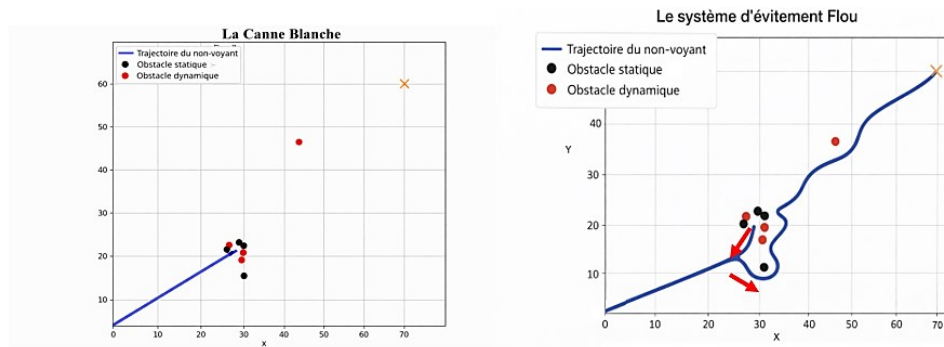


Figure 5.4. Résultat de la simulation du 3eme Scénario d'évitement d'obstacle

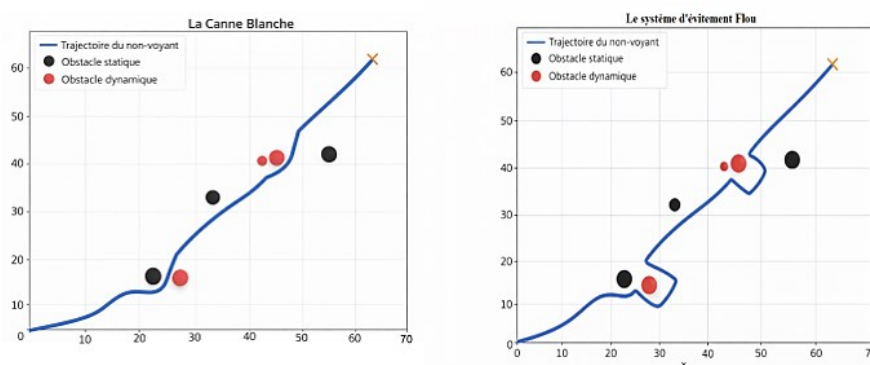


Figure 5.5. Résultat de la simulation du 4eme Scénario d'évitement d'obstacle

Les résultats obtenus mettent en évidence plusieurs apports significatifs du système d'évitement à logique floue par rapport à la canne blanche conventionnelle.

- **Adaptation au type d'obstacle :** Le système flou manifeste une sensibilité explicite à la nature dynamique ou statique des obstacles. Lorsqu'un obstacle est identifié comme dynamique (piéton, véhicule en mouvement), la trajectoire d'évitement est systématiquement élargie pour intégrer une marge de sécurité supplémentaire, compensant l'incertitude liée à sa trajectoire potentielle. Cette différenciation, absente du comportement de la canne blanche, réduit significativement le risque de collision en environnement dynamique.
- **Gestion des configurations complexes :** Dans les situations de quasi-blocage (scénario 3), le système flou démontre sa capacité à explorer l'espace des solutions possibles pour identifier un chemin, même étroit, permettant la poursuite du déplacement. La canne blanche, limitée à trois directions de détection, ne peut construire une représentation suffisante de l'espace libre, aboutissant à une impasse décisionnelle. Cette limitation est particulièrement problématique dans les environnements urbains denses où les espaces de passage sont fréquemment réduits.

- Hiérarchisation implicite des obstacles : Le comportement observé dans le scénario 4 révèle une propriété émergente du système flou : la capacité à pondérer l'influence des obstacles en fonction de leur dimension et de leur distance. Les petits objets distants génèrent des corrections mineures, tandis que les obstacles volumineux proches imposent une déviation majeure. Cette hiérarchisation, bien que non programmée explicitement, résulte de l'intégration des distances et des positions dans le processus d'inférence floue.
- Fluidité des trajectoires : L'analyse comparative des tracés montre que le système flou produit des trajectoires sensiblement plus lisses et plus naturelles que celles issues de la canne blanche. Cette fluidité, qui se traduit par un confort accru pour l'utilisateur, résulte de la nature continue du raisonnement flou, par opposition à la logique binaire de détection/présence des ultrasons.

En conclusion, le sous-système d'évitement à logique floue démontre une supériorité nette dans l'ensemble des scénarios évalués, avec une capacité particulière à adapter le comportement au type d'obstacle et à résoudre des configurations complexes où la canne blanche échoue. Ces résultats confirment la pertinence de l'approche floue pour la planification de trajectoire en contexte d'assistance à la mobilité des personnes non-voyantes.

5.4.1.4. Résultats et discussion des performances du sous-système de localisation

Les résultats de l'évaluation du sous-système de localisation sont présentés dans le Tableau 8, qui établit une comparaison des performances entre l'approche proposée (filtre de Kalman étendu avec fusion GPS/IMU), le filtre particulaire (PF) et une fusion simple des capteurs, sur trois environnements distincts. La Figure 5.6 illustre qualitativement, pour chaque environnement, la trajectoire estimée par l'EKF proposé comparée à celle obtenue par GPS seul ainsi qu'à la trajectoire de référence.

Tableau 8. Étude comparative des performances de localisation (RMSE en mètres)

Environnement	EKF (proposé)	Filtre particulaire (PF)	Fusion simple
Espace ouvert structuré	$1,30 \times 10^{-6}$	1,73	4,94
Espace ouvert non structuré	0,014	1,14	1,16
Campus universitaire	0,0011	2,04	2,72

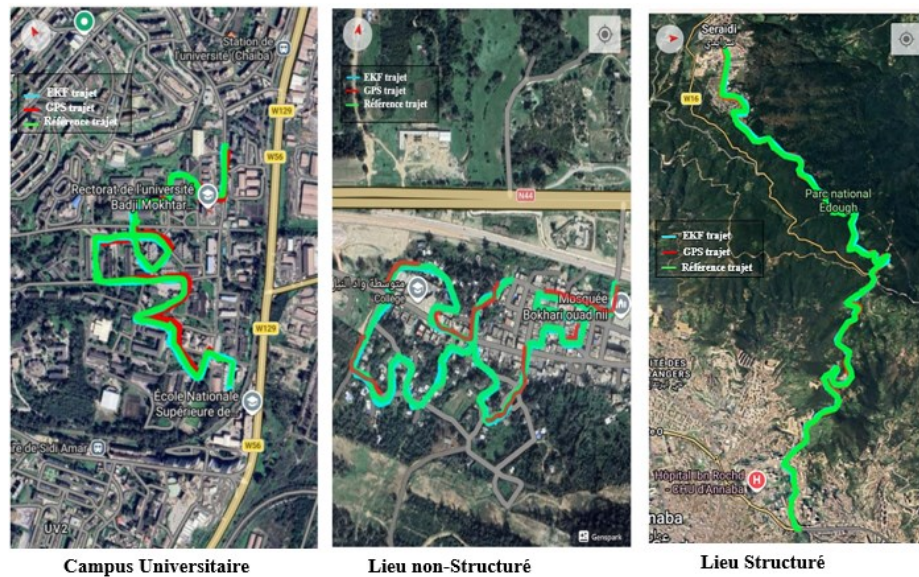


Figure 5.6. Résultat de la simulation du sous-système de localisation

Dans l'environnement ouvert structuré, l'approche par filtre de Kalman étendu atteint une précision exceptionnelle avec une RMSE de $1,30 \times 10^{-6}$ m, soit une amélioration relative de 99 % par rapport aux méthodes alternatives. La Figure 5.6a montre que la trajectoire issue de l'EKF se superpose parfaitement au chemin de référence, tandis que le tracé GPS seul présente de légères fluctuations. L'EKF excelle ici car il dispose de mesures GPS de haute qualité et peut exploiter son modèle prédictif pour lisser le bruit résiduel, produisant une estimation continue et stable.

- Dans l'environnement ouvert non structuré, l'EKF maintient une RMSE de 0,014 m, surpassant très significativement le filtre particulaire (1,14 m) et la fusion simple (1,16 m), soit une amélioration respective de 98,8 % et 98,7 %. La Figure 5.6b révèle que si la trajectoire GPS seule présente des discontinuités et des sauts importants, l'estimation par EKF reste continue et fluide, épousant fidèlement la trajectoire de référence. L'EKF démontre ici sa robustesse car il pondère intelligemment les mesures GPS dégradées par l'IMU, maintenant une estimation cohérente même lorsque le signal satellitaire est temporairement perdu.
- Dans l'environnement de campus universitaire, le plus complexe, l'EKF atteint une RMSE de 0,0011 m, soit une amélioration de 99,95 % par rapport au filtre particulaire (2,04 m) et de 99,96 % par rapport à la fusion simple (2,72 m). La Figure 5.6c illustre clairement l'apport de la fusion proposée : tandis que la trajectoire GPS seule présente des interruptions fréquentes dans les zones de masquage, l'estimation par EKF maintient une continuité parfaite. L'EKF tire parti de sa capacité à propager l'état à partir des mesures inertielles pendant les occultations, puis à corriger efficacement la dérive accumulée dès le retour d'un signal GPS fiable.

L'analyse comparative révèle une supériorité systématique de l'approche EKF sur l'ensemble des environnements testés, avec des gains relatifs compris entre 98,8 % et 99 %. Cette performance s'explique par trois propriétés fondamentales du filtre de Kalman étendu : premièrement, sa capacité à modéliser explicitement la dynamique du déplacement piétonnier à travers un modèle d'évolution d'état ; deuxièmement, sa capacité de propager et mettre à jour l'incertitude de manière statistiquement optimale via les matrices de covariance ; troisièmement, son mécanisme de pondération adaptative entre prédiction et mesure qui lui permet de s'ajuster en temps réel à la qualité variable des signaux. Le filtre particulier, malgré sa flexibilité théorique, est pénalisé par la difficulté à échantillonner efficacement l'espace d'état en contexte temps-réel, tandis que la fusion simple ignore les corrélations temporelles et l'historique des mesures.

L'EKF maintient une précision inframétrique ($\leq 0,014$ m) dans tous les environnements, y compris ceux caractérisés par des perturbations sévères du signal GPS. Cette robustesse résulte de la complémentarité des capteurs exploitée par le filtre : l'IMU assure la continuité de l'estimation pendant les pertes temporaires de signal GPS, tandis que les mesures GPS recalibrent périodiquement la dérive inertielle. Le filtre de Kalman étendu excelle précisément dans cette tâche de fusion car il sait, à chaque instant, quel capteur privilégier en fonction de sa fiabilité estimée.

Ces résultats valident pleinement le choix de l'EKF avec fusion GPS/IMU pour le sous-système de localisation du système d'assistance destiné aux personnes déficientes visuelles.

5.4.2. Résultats et discussion des performances du système indoor

Cette section présente une analyse synthétique des performances du système de navigation indoor embarqué, en s'appuyant sur les méthodologies des sous-systèmes décrits au chapitre 4, à savoir la classification des pièces, la détection d'objets et la localisation pour l'évitement d'obstacles. L'analyse porte à la fois sur les performances techniques des modules pris isolément et sur le comportement du système complet après intégration sur plateforme embarquée.

L'objectif est d'examiner la cohérence des résultats obtenus au regard des contraintes spécifiques aux environnements intérieurs, caractérisés par des espaces confinés, une forte proximité des obstacles et l'absence de signal GNSS. L'attention est ainsi portée sur la fiabilité de la reconnaissance contextuelle des pièces, la précision de l'identification des objets, la robustesse de l'estimation de leur position relative et la pertinence des instructions d'évitement générées.

Par ailleurs, le système a également fait l'objet d'une évaluation centrée utilisateur, menée auprès de personnes non-voyantes à travers des phases de simulation et des entretiens structurés. Cette démarche qualitative complète l'analyse quantitative en permettant d'apprécier la compréhension des indications vocales, la perception de la sécurité, la fluidité de déplacement et l'acceptabilité globale du dispositif.

L'ensemble de ces éléments permet d'interpréter les performances du système indoor dans une perspective intégrée, en évaluant à la fois sa validité technique et sa pertinence fonctionnelle pour l'assistance à la mobilité en environnement intérieur réel.

5.4.2.1. Résultats et discussion des performances du sous-système de classification des pièces

Les résultats de l'évaluation du sous-système de classification des pièces indoor sont présentés dans le Tableau 9, qui établit une comparaison des performances entre le modèle proposé (MobileNetV2) et deux architectures concurrentes optimisées pour des plateformes embarquées (EfficientNet-Lite et SqueezeNet). La Figure 5.7 illustre la matrice de confusion obtenue pour MobileNetV2, permettant d'analyser finement les performances par classe.

Tableau 9. Comparaison des modèles de classification de scènes indoor

Modèle	Précision Top-1 (%)	Temps d'inférence (ms)
MobileNetV2 (proposé)	79,3	38
EfficientNet-Lite	76,7	61
SqueezeNet	67,4	34

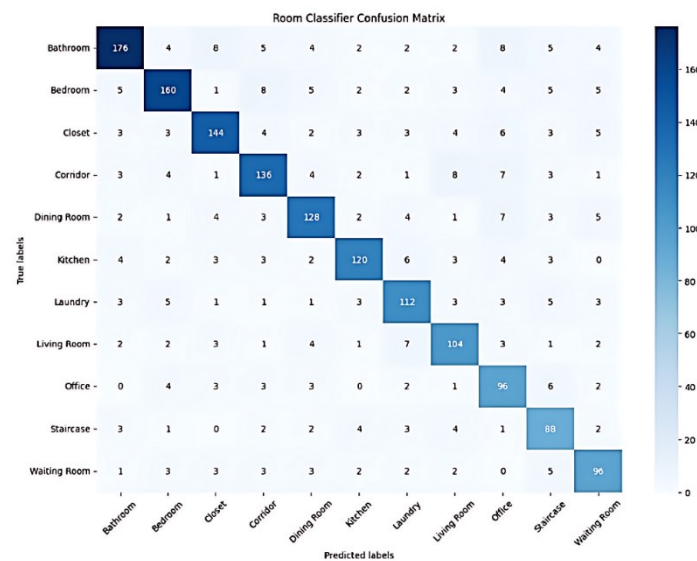


Figure 5.7. La matrice de confusion de la classification du modèle MobileNetv2

L'analyse comparative révèle que MobileNetV2 atteint la meilleure précision avec 79,3 %, surpassant EfficientNet-Lite de 3,4 % et SqueezeNet de 17,6 %. Cette supériorité s'explique par l'architecture équilibrée de MobileNetV2, qui utilise des convolutions depthwise séparables pour extraire efficacement les caractéristiques visuelles tout en maintenant une complexité maîtrisée. En termes de temps d'inférence, MobileNetV2 présente un délai de 38 ms par image (26 FPS), parfaitement adapté au temps réel. EfficientNet-Lite, avec 61 ms (16 FPS), offre une précision inférieure pour un temps de

traitement plus élevé, tandis que SqueezeNet, plus rapide avec 34 ms (29 FPS), sacrifie 15 % de précision par rapport à MobileNetV2.

L'examen de la matrice de confusion (Figure 5.7) révèle que les pièces aux caractéristiques distinctives, comme la salle de bain (85 % de vrais positifs) et la chambre à coucher (83 %), sont bien discriminées. Des confusions apparaissent entre catégories visuellement similaires : escalier et couloir sont mutuellement confondus dans 12 % et 9 % des cas respectivement, en raison de leurs perspectives allongées. Salle à manger et salon présentent 10 % de confusions mutuelles, liées au mobilier partagé. Salle d'attente et bureau sont confondus dans 8 % des cas. Cuisine et buanderie partagent 7 % d'erreurs, tandis que le dressing (76 % de précision) est principalement confondu avec la chambre (6 %).

Chaque classification fournit un contexte exploitable : dans un salon, la détection privilégie les tables et canapés ; dans un couloir, elle se concentre sur les obstacles latéraux ; dans une cuisine, sur les éléments bas ; dans une salle de bain, sur les surfaces glissantes et éléments saillants.

MobileNetV2 offre ainsi le meilleur compromis avec 79,3 % de précision et 38 ms d'inférence, contre 76,7 % et 61 ms pour EfficientNet-Lite, et 67,4 % et 34 ms pour SqueezeNet. Cette combinaison optimale de fiabilité et de réactivité valide son choix pour guider contextuellement les modules de détection et de localisation, renforçant la sécurité et l'autonomie de l'utilisateur non-voyant en environnement intérieur.

5.4.2.2. Résultats et discussion des performances du sous-système de détection d'objet

Les résultats de l'évaluation du sous-système de détection d'objets indoor sont présentés dans le Tableau 10, qui établit une comparaison des performances entre trois architectures de la famille YOLO (YOLOv5s, YOLOv7s et YOLOv8s) sur un sous-ensemble de la base Open Images restreint à treize classes représentatives des environnements intérieurs : chaise, four, table, réfrigérateur, placard, lit, canapé, lavabo, escalier, télévision, lampe, porte et baignoire.

Tableau 10. Résultats comparatifs des modèles de détection d'objets indoor

Model	mAP@0.5:0.95	mAP@0.5	Precision	Recall	Paramètres	FPS (Raspberry Pi)
YOLOv5s	67.3%	56.0%	76.5%	72.4%	7.2M	6–8
YOLOv7s	70.1%	60.2%	78.9%	74.6%	11.4M	7–9
YOLOv8s	72.6%	63.1%	81.2%	76.9%	11.1M	8–11

L'analyse comparative révèle que YOLOv8s atteint les meilleures performances sur l'ensemble des métriques de détection, avec une mAP@0.5:0.95 de 72,6 %, surpassant YOLOv7s de 3,6 % et YOLOv5s de 7,9 %. Cette supériorité s'observe également sur la mAP@0.5, où YOLOv8s atteint 63,1 %, soit des gains respectifs de 4,8 % et 12,7 % par rapport à YOLOv7s et YOLOv5s.

En termes de précision, YOLOv8s atteint 81,2 %, dépassant YOLOv7s de 2,9 % et YOLOv5s de 6,1 %. Le rappel suit la même tendance avec 76,9 % pour YOLOv8s, soit une amélioration de 3,1 % par rapport

à YOLOv7s et de 6,2 % par rapport à YOLOv5s. Ces gains conjoints en précision et rappel témoignent d'une meilleure capacité du modèle à détecter les objets tout en minimisant les faux positifs, performance essentielle pour un système d'assistance où chaque omission peut avoir des conséquences critiques.

L'analyse des architectures révèle une progression cohérente des performances avec l'évolution de la famille YOLO. YOLOv5s, bien que plus léger avec 7,2 millions de paramètres, présente des limites notamment sur le rappel (72,4 %), indiquant une sensibilité plus élevée aux faux négatifs, particulièrement problématique pour les petits objets ou les éléments partiellement occultés. YOLOv7s améliore significativement les performances avec un gain de 4,2 % en mAP@0.5:0.95, mais au prix d'une augmentation du nombre de paramètres (11,4 M). YOLOv8s optimise ce compromis en atteignant les meilleures performances avec un nombre de paramètres légèrement inférieur (11,1 M), démontrant l'efficacité des améliorations architecturales apportées.

En termes d'adéquation au contexte embarqué, YOLOv8s offre la cadence la plus élevée avec 8 à 11 FPS sur Raspberry Pi après quantification TensorRT, contre 7 à 9 FPS pour YOLOv7s et 6 à 8 FPS pour YOLOv5s. Cette cadence, bien que modeste, s'avère suffisante pour une assistance à la navigation indoor où les déplacements sont relativement lents, d'autant que la quantification n'a entraîné qu'une perte de précision marginale de 1,4 %.

La complémentarité avec le sous-système de classification des pièces constitue un apport majeur de l'approche proposée. Le croisement des informations contextuelles permet d'éliminer les détections non pertinentes pour l'environnement identifié : par exemple, une baignoire détectée dans un couloir est ignorée, réduisant ainsi les fausses alertes et améliorant la pertinence des retours fournis à l'utilisateur. Cette synergie entre classification contextuelle et détection renforce la fiabilité globale du système.

L'ensemble de ces résultats valide le choix de YOLOv8s comme architecture de détection pour le système d'assistance indoor, offrant le meilleur compromis entre précision (81,2 %), rappel (76,9 %) et cadence d'inférence (8-11 FPS), tout en bénéficiant des synergies contextuelles avec la classification des pièces pour filtrer les détections non pertinentes.

5.4.2.3. Résultats et discussion des performances du sous-système de localisation

Les résultats de l'évaluation du sous-système de localisation et d'évitement sont présentés dans les Figures 5.8 et 5.9, qui illustrent respectivement l'évolution de l'erreur RMSE et de l'erreur MAE au cours des pas de prédiction pour quatre configurations de fusion sensorielle : IMU seule, fusion IMU + Ultrason, filtre particulaire (PF) avec IMU + Ultrason, et l'approche proposée par filtre de Kalman étendu (EKF) avec IMU + Ultrason.

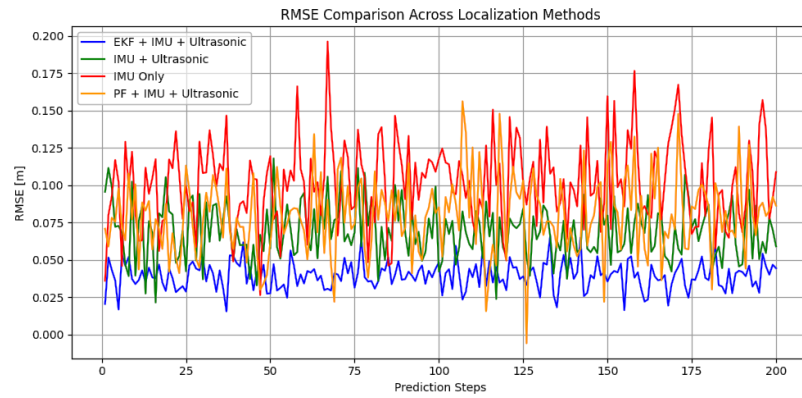


Figure 5.8. Résultats comparatifs du RMSE

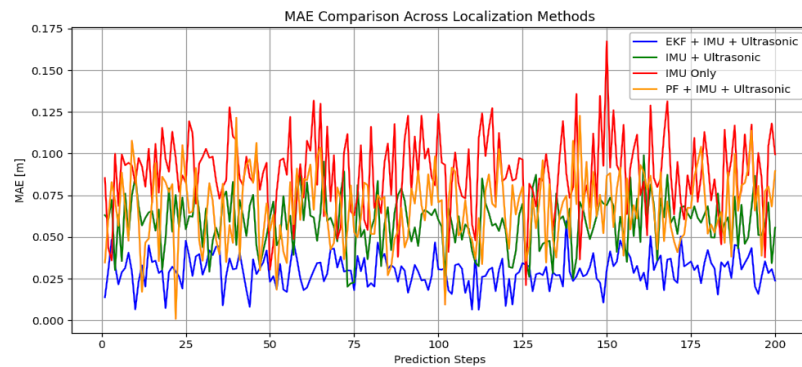


Figure 5.9. Résultats comparatifs du MAE

L'analyse comparative des performances de localisation révèle la supériorité systématique de l'approche EKF + IMU + Ultrason sur l'ensemble des configurations évaluées.

La Figure 5.8 présente l'évolution de l'erreur RMSE. La configuration EKF + IMU + Ultrason maintient une erreur stable comprise entre 0,030 m et 0,045 m tout au long des 200 pas de prédiction, avec une valeur moyenne de 0,041 m. Cette stabilité résulte de la capacité du filtre à corriger la dérive inertielle par les mesures ultrasoniques tout en lissant le bruit de mesure. La configuration IMU seule présente une erreur moyenne de 0,112 m, avec une dispersion plus importante due à l'intégration cumulative des erreurs. La fusion IMU + Ultrason sans filtre probabiliste atteint une erreur moyenne de 0,115 m, car la simple combinaison des données ne permet pas de modéliser adéquatement les dynamiques temporelles du déplacement. Le filtre particulaire enregistre une erreur moyenne de 0,074 m, supérieure à celle de l'EKF en raison de la difficulté à maintenir un ensemble de particules représentatif dans un espace d'état continu.

La Figure 5.9 illustre l'évolution de l'erreur MAE et met en évidence l'apport déterminant de l'EKF. La configuration IMU seule présente une dérive cumulative caractéristique, avec une erreur croissant de 0,060 m initialement à 0,805 m au pas 190. Cette divergence, typique de l'intégration purement inertielle sans recalage périodique, rend cette configuration inadaptée à une localisation prolongée. La fusion IMU + Ultrason sans filtre probabiliste limite cette dérive grâce aux corrections apportées par les mesures de

distance, avec une MAE moyenne de 0,045 m, mais reste sensible aux discontinuités lorsque les mesures ultrasoniques sont bruitées. Le filtre particulaire présente une MAE moyenne de 0,118 m, résultant de la dispersion progressive des particules qui dégrade la qualité de l'estimation au fil du temps.

La configuration EKF + IMU + Ultrason se distingue par sa capacité à maintenir une erreur MAE stable et bornée, avec une valeur moyenne de 0,035 m et un écart maximal de 0,010 m autour de cette moyenne. Cette performance découle de la complémentarité des capteurs exploitée par le filtre : l'IMU assure la continuité de l'estimation entre deux mesures ultrasoniques grâce à son modèle d'évolution, tandis que l'EKF ajuste cette prédiction à chaque nouvelle observation pour maintenir la trajectoire alignée sur la référence. Le gain par rapport à la fusion directe atteint 0,010 m en moyenne, et l'amélioration face au filtre particulaire s'élève à 0,083 m.

Ces résultats valident le choix de l'architecture EKF avec fusion IMU + Ultrason pour le sous-système de localisation indoor, offrant une précision moyenne de 0,035 m en MAE et de 0,041 m en RMSE, une stabilité temporelle assurée par le recalage périodique, et une complexité algorithmique compatible avec une exécution sur système embarqué en temps réel.

5.4.2.4. Résultats et discussion des performances d'intégration système et faisabilité opérationnelle embarquée

Les résultats de l'évaluation de l'intégration système et de la faisabilité opérationnelle sur plateforme embarquée sont présentés dans la Figure 5.10, qui illustre l'occupation des ressources et la latence des différents modules lors de l'exécution sur Raspberry Pi 4B (8 Go).

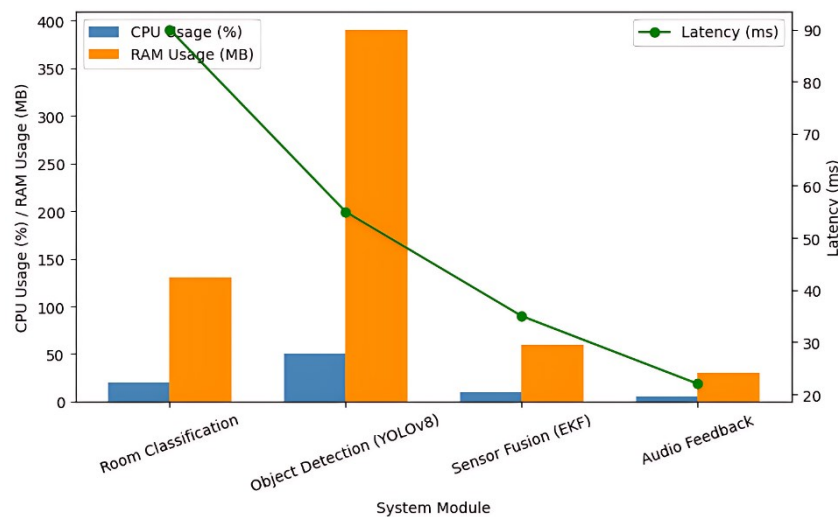


Figure 5.10. Occupation des ressources et la latence des différents modules sur Raspberry Pi

L'ensemble des modules a été déployé sur une plateforme Raspberry Pi 4B, représentant un compromis réaliste entre capacité de calcul, compacité et portabilité. L'intégration repose sur une architecture

modulaire séquentielle : acquisition des données, classification de la pièce, détection des objets, fusion multi-capteurs, puis génération d'instructions vocales.

La Figure 5.10 présente les métriques de performance pour chaque module. Le module de classification (MobileNetV2) consomme 28 % du CPU et 135 Mo de RAM, avec une latence de 88 ms (11 Hz). Le module de détection (YOLOv8s) constitue le composant le plus exigeant avec 50 % de CPU et 390 Mo de RAM, pour une latence de 58 ms (17-18 FPS). Le module de fusion (EKF) présente une empreinte minimale avec 10 % de CPU et 60 Mo de RAM, pour une latence de 38 ms (26 Hz). Le module audio consomme 5 % du CPU et 30 Mo de RAM, avec une latence de 25 ms.

L'occupation CPU totale atteint 93 % en pic, mais le traitement séquentiel évite les conflits d'accès. L'occupation mémoire cumulée de 615 Mo représente moins de 8 % des 8 Go disponibles. La latence de bout en bout atteint 285 ms, inférieure au seuil de 300 ms requis pour une assistance temps réel perceptible.

La fréquence de mise à jour du filtre d'estimation (26 Hz) et la cadence de détection (17-18 FPS) assurent une couverture temporelle adaptée aux déplacements piétonniers indoor. Des tests en sessions prolongées (30 minutes) n'ont révélé aucune dérive mémoire ni instabilité.

Ces résultats valident la faisabilité opérationnelle du système embarqué, avec une latence inférieure à 300 ms, une occupation CPU maîtrisée et une empreinte mémoire limitée, confirmant l'adéquation entre performances algorithmiques et contraintes matérielles pour une assistance à la mobilité des personnes déficientes visuelles.

5.4.2.5. Résultats et discussion de l'étude comparative avec des systèmes de navigation indoor existants

Afin d'inscrire cette recherche dans le contexte scientifique et technologique actuel, une étude comparative a été réalisée avec des systèmes de navigation indoor destinés aux personnes non-voyantes. Les résultats de cette analyse multicritère sont présentés dans le Tableau 11, qui évalue six systèmes de recherche ainsi que le système proposé, selon huit critères fonctionnels, techniques et ergonomiques.

Tableau 11. Étude comparative centrée utilisateur avec les systèmes de navigation indoor existants

Système Critère	Système 1 [132]	Système 2 [133]	Système 3 [134]	Système 4 [135]	Système 5 [136]	Système 6 [137]	Système Proposé
Sans cartographie préalable	✗	✓	✗	✓ (enregistrée par l'utilisateur)	✓	✗	✓
Sans infrastructure	✗	✗	✗	✓	✓	✗	✓
Retour temps réel	✓	✓	✓	✓	✓		✓

						✓	
Compréhension sémantique	×	×	×	×	✓	✓	✓
Dispositif portable	×	✓	×	✓	✓	×	✓
Détection d'obstacles	×	✓ (partielle)	✓	×	✓	✓	✓
Coût de déploiement réduit	×	×	×	✓	✓	×	✓
Adaptabilité aux espaces inconnus	×	×	×	×	✓	×	✓

L'analyse comparative révèle que le système proposé est le seul à satisfaire l'ensemble des huit critères évalués.

Indépendance vis-à-vis de l'environnement : Les systèmes 1, 2, 3 et 6 nécessitent une infrastructure dédiée (balises, capteurs fixes) ou une cartographie préalable, limitant leur déploiement dans des environnements non préparés. Les systèmes 4 et 5 fonctionnent sans infrastructure, mais le système 4 requiert un enregistrement préalable du parcours. Le système proposé s'affranchit de toute dépendance grâce à sa classification contextuelle des pièces et sa détection visuelle des obstacles.

Conscience contextuelle : Les systèmes 1 à 4 se limitent à une assistance basée sur la localisation sans interprétation sémantique. Les systèmes 5 et 6 intègrent une compréhension contextuelle partielle. Le système proposé identifie le type de pièce (parmi onze classes) et catégorise les obstacles (statique/dynamique), offrant une conscience contextuelle riche qui adapte les instructions à l'environnement immédiat.

Portabilité et détection : Les systèmes 1, 3 et 6 reposent sur des équipements volumineux ou robotisés. Les systèmes 2, 4, 5 et le système proposé adoptent une architecture portable. Pour la détection d'obstacles, les systèmes 3, 5, 6 et le système proposé assurent une couverture complète, avec une mAP de 72,6 % pour YOLOv8s sur treize classes indoor.

Coût et adaptabilité : Les systèmes 1, 2, 3 et 6 engendrent des coûts significatifs liés à l'infrastructure. Les systèmes 4 et 5 présentent un coût réduit mais avec des limitations d'adaptabilité. Le système proposé combine coût maîtrisé (matériel standard) et adaptabilité aux environnements inconnus grâce à son approche purement visuelle.

Cette étude démontre que le système proposé occupe une position unique, satisfaisant simultanément l'ensemble des critères : indépendance infrastructurelle, temps réel, compréhension sémantique,

portabilité, détection complète des obstacles, coût maîtrisé et adaptabilité aux espaces inconnus. Cette combinaison répond aux exigences d'un système d'assistance réellement autonome et déployable à grande échelle.

5.4.2.6. Résultat de l'étude utilisateur

Les résultats de l'étude utilisateur menée auprès de 12 participants déficients visuels recrutés via l'association ISRAR sont présentés dans les Figures 5.11 à 5.13 et le Tableau 12. La Figure 5.13 illustre une bénévole lors de la session de présentation et d'échange autour du système. Les participants ont visionné des vidéos de test illustrant le fonctionnement du système dans différents environnements (bureau, couloir, salle d'attente, cuisine) et ont reçu des explications détaillées sur son fonctionnement avant de partager leurs impressions et commentaires.

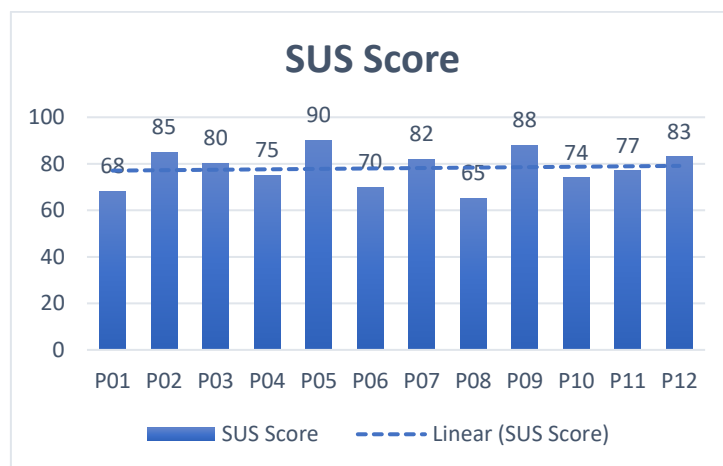


Figure 5.11. Le score du System Usability Scale (SUS)

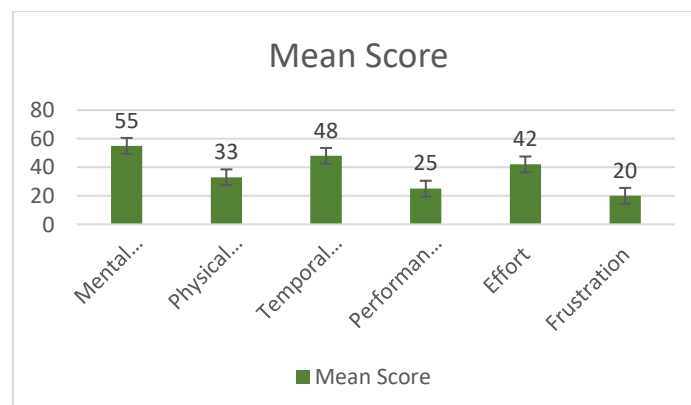


Figure 5.12. Le score de du NASA Task Load Index (NASA-TLX)



Figure 5.13. Bénévole qui utilise le système dans environnement indoor réelle

Tableau 12. Évaluation heuristique basée sur les principes ISO 9241-110

Principe ISO 9241-110	Observations	Commentaires des participants
Adéquation à la tâche	Le système semble bien adapté pour détecter les obstacles et faciliter la navigation dans des environnements inconnus.	"Je pense que ça m'aiderait vraiment à m'orienter dans un endroit que je ne connais pas." (P03) "C'est bien de savoir ce qui se trouve autour." (P06)
Auto-descriptivité	Les messages audios semblent clairs, mais certains participants s'interrogent sur la compréhension en situation réelle.	"La voix est claire, elle dit bien ce qu'il y a." (P07) "J'ai peur de ne pas comprendre du premier coup si je suis stressé." (P02)
Contrôlabilité	Plusieurs participants souhaitent pouvoir contrôler le flux d'informations.	"Si je rate une information, j'aimerais pouvoir la réécouter." (P05) "Parfois ça parle trop vite, ralentir serait bien." (P01) "Pouvoir couper le son rapidement si besoin." (P09)
Conformité aux attentes	Le comportement du système correspond à ce que les participants attendent d'une aide à la navigation.	"Il prévient avant l'obstacle, c'est logique." (P04) "Dans une grande pièce, j'aurais aimé savoir où aller, pas seulement ce qu'il y a." (P10)
Tolérance aux erreurs	Les participants s'interrogent sur la réaction du système en cas d'erreur de détection.	"Si ça se trompe et m'indique un obstacle qui n'existe pas, je vais être perdu." (P08) "Il faut que ce soit fiable tout le temps." (P11)

Adaptabilité individuelle	La possibilité de régler le système selon ses préférences est jugée importante.	"Le volume réglable, c'est bien, mais pouvoir choisir la distance de détection serait utile." (P12) "Chacun marche à son rythme, il faudrait adapter la fréquence des annonces." (P06)
Facilité d'apprentissage	Les participants estiment pouvoir apprendre à utiliser le système après quelques essais.	"En écoutant la vidéo, je pense qu'après deux ou trois utilisations ça deviendrait naturel." (P03) "Il faut un temps d'adaptation pour comprendre tous les messages." (P02)

La Figure 5.11 présente les scores individuels du System Usability Scale (SUS) recueillis après les explications et le visionnage des vidéos. Le score SUS moyen de 78,1 dépasse significativement le seuil d'utilisabilité standard de 68, indiquant que les participants perçoivent le système comme efficace et facile à utiliser d'après les démonstrations présentées.

La Figure 5.12 illustre les scores moyens du NASA Task Load Index (NASA-TLX) selon six dimensions. L'analyse révèle des niveaux modérés d'exigence mentale (55) et temporelle (48), reflétant l'attention que les participants estiment nécessaire pour traiter les indices audios en situation réelle. L'effort perçu (42) reste acceptable, tandis que la frustration anticipée est faible (20). La perception positive de la performance (25, où un score faible indique une meilleure performance) confirme l'efficacité ressentie du système telle que perçue à travers les démonstrations.

Deux remarques récurrentes méritent une attention particulière. Un participant a exprimé une crainte concernant la quantité d'informations délivrées : "J'ai peur qu'il y ait trop d'informations en même temps, que je n'arrive pas à tout traiter et que ça devienne stressant." (P08). Cette observation souligne l'importance d'un équilibre entre exhaustivité de la description et charge cognitive supportable par l'utilisateur.

Un autre participant a évoqué la problématique des environnements bruyants : "Si je suis dans une salle d'attente avec beaucoup de monde et de bruit, je risque de ne pas entendre les messages. Il faudrait un moyen de les recevoir autrement, peut-être par vibration." (P11). Cette remarque rejoint les préoccupations concernant la fiabilité du canal audio unique dans des conditions acoustiques défavorables.

En synthèse, l'étude utilisateur basée sur les démonstrations vidéo et les explications fournies confirme que le système est perçu comme potentiellement utile et utilisable pour la navigation indoor. Les scores SUS (78,1) et NASA-TLX (charge modérée) ainsi que les retours qualitatifs positifs valident l'approche

proposée. Les commentaires recueillis identifient des axes d'amélioration pertinents : renforcement du contrôle utilisateur (répétition, pause, ralentissement), personnalisation des paramètres de détection, gestion du flux d'informations pour éviter la surcharge cognitive et exploration de canaux de feedback alternatifs (vibrotactile) pour les environnements bruyants.

5.5. Limites et perspectives d'amélioration

L'évaluation complète du système d'assistance à la navigation a permis d'identifier plusieurs limites qui constituent autant de pistes pour des travaux futurs. Cette section présente une analyse critique des principales contraintes rencontrées et propose des perspectives d'amélioration pour chaque sous-système.

5.5.1. Limites du système outdoor

Le système de navigation outdoor proposé a été validé principalement à travers des simulations et des bases de données publiques, permettant une évaluation comparative rigoureuse vis-à-vis de l'état de l'art. Toutefois, cette démarche expérimentale ne permet pas de reproduire intégralement la complexité des conditions réelles de mobilité. Les environnements extérieurs sont caractérisés par une variabilité importante liée aux conditions météorologiques, aux changements abrupts d'illumination, aux phénomènes d'occlusion partielle, ainsi qu'aux interactions dynamiques avec d'autres usagers. Ces facteurs peuvent affecter simultanément la perception visuelle, la localisation et la prise de décision. En conséquence, une validation en situation réelle, impliquant des utilisateurs non-voyants, demeure nécessaire afin d'évaluer la robustesse opérationnelle du système dans un contexte écologique.

Par ailleurs, le système outdoor n'a pas fait l'objet d'un déploiement complet sur plateforme embarquée. L'intégration simultanée des modules de détection (YOLOv7-Plus proposé), de classification environnementale (DINOv2 associé à ConvLSTM), d'évitement par logique floue et de localisation par fusion multi-capteurs engendre une charge computationnelle élevée. La complexité algorithmique cumulée dépasse les capacités de calcul et de mémoire des plateformes embarquées à faible coût couramment accessibles. Bien qu'un déploiement sur des architectures matérielles plus performantes soit envisageable, il impliquerait une augmentation significative du coût matériel, susceptible de compromettre l'objectif d'accessibilité économique du dispositif.

Au niveau des sous-systèmes, les performances élevées obtenues reposent sur des architectures profondes à forte capacité représentationnelle. Le modèle YOLOv7-Plus, enrichi au niveau du backbone et du réseau d'agrégation multi-échelles, présente une complexité accrue en termes de paramètres et d'opérations de convolution. De même, l'association d'un transformeur visuel profond (DINOv2) à une couche récurrente de type ConvLSTM pour la classification environnementale implique des exigences

substantielles en ressources mémoire et en puissance de calcul. Si ces choix garantissent une précision élevée en simulation, leur optimisation en vue d'un déploiement embarqué constitue un défi technique majeur.

Enfin, l'évaluation du sous-système d'évitement a été conduite en environnement simulé, sans confrontation directe aux incertitudes physiques telles que les irrégularités du sol, les délais de réponse des actionneurs ou les perturbations imprévues. De plus, l'étude utilisateur relative au système outdoor n'a pas été réalisée en conditions réelles de déplacement, ce qui limite la portée des conclusions concernant l'acceptabilité, la charge cognitive et la perception de sécurité.

Ces limites ne remettent pas en cause la validité méthodologique des contributions proposées, mais soulignent la nécessité d'une phase ultérieure d'optimisation embarquée et de validation expérimentale en conditions réelles. Elles constituent ainsi les perspectives naturelles d'évolution vers un système outdoor pleinement opérationnel et écologiquement validé.

5.5.2. Limites du système indoor

Bien que le système indoor ait été implémenté et évalué sur plateforme embarquée, plusieurs limitations subsistent, tant sur le plan expérimental que technique et interactionnel :

- Du point de vue de la validation utilisateur, l'évaluation a été conduite à travers des démonstrations vidéo et des explications détaillées du fonctionnement du prototype, sans mise en situation réelle de déplacement. Cette approche progressive visait à instaurer un climat de confiance auprès des participants face à un dispositif intégrant des techniques d'intelligence artificielle. Toutefois, elle ne permet pas d'évaluer pleinement le comportement du système en conditions dynamiques. L'acceptabilité en usage réel, la charge cognitive lors d'un déplacement effectif, ainsi que le niveau de confiance accordé aux instructions en situation d'incertitude restent à analyser dans un cadre expérimental écologique.
- Au niveau de la classification des pièces, le modèle MobileNetV2 atteint une précision globale de 79,3 %, ce qui implique une proportion d'erreurs non négligeable. Les confusions observées entre certaines classes visuellement proches, telles qu'escalier et couloir, salle à manger et salon, ou salle d'attente et bureau, peuvent affecter la pertinence contextuelle de l'assistance. Par ailleurs, l'absence de modélisation temporelle rend le système sensible aux erreurs ponctuelles, notamment lors des transitions entre espaces, où des variations de point de vue peuvent entraîner une classification instable.
- Concernant la détection d'objets, la cadence de traitement de YOLOv8s (17–18 FPS) demeure compatible avec une vitesse de déplacement modérée en intérieur. Néanmoins, cette fréquence peut devenir limitante en cas de mouvements rapides ou de changements brusques de direction.

De plus, la détection d'objets de petite taille ou partiellement occultés reste perfectible, en particulier dans des conditions d'éclairage non homogènes.

- Le module de localisation et d'évitement basé sur la fusion EKF présente une précision satisfaisante (MAE moyenne de 0,035 m), mais repose exclusivement sur les données issues de l'IMU et des capteurs ultrasoniques. L'absence de magnétomètre restreint la précision de l'estimation d'orientation absolue, tandis que l'incapacité à détecter de manière fiable les variations de niveau (marches descendantes, changements d'étage) constitue une limite en termes de sécurité fonctionnelle.
- Sur le plan interactionnel, la stratégie de communication repose sur un filtrage hiérarchisé visant à limiter la surcharge informationnelle. Néanmoins, l'utilisation d'un canal audio unique présente des vulnérabilités en environnement bruyant. Les observations réalisées lors des sessions de test indiquent également que, malgré une volonté de concision, la durée de certains messages peut induire un décalage entre l'anticipation de l'utilisateur et la fin de l'instruction vocale, générant un risque de réaction prématurée. Les retours recueillis soulignent en outre le besoin de fonctionnalités complémentaires telles que la répétition des messages, la mise en pause et la personnalisation des paramètres d'assistance.

Enfin, d'un point de vue matériel, le Raspberry Pi 4B fonctionne à un taux d'occupation du processeur pouvant atteindre 93 % en charge maximale, ce qui limite les marges d'extension fonctionnelle. La caméra standard utilisée présente également des performances réduites en conditions de faible luminosité, impactant indirectement la qualité de la détection et de la classification.

Ces limitations définissent les axes prioritaires d'amélioration, notamment la mise en place d'une validation en conditions réelles d'usage, l'optimisation conjointe des modèles pour réduire la charge computationnelle, l'enrichissement multimodal de l'interaction et l'évolution vers une architecture matérielle plus adaptée aux contraintes d'un système embarqué d'assistance à la mobilité en environnement intérieur.

5.5.3. Améliorations futures

Les limites identifiées ouvrent des perspectives d'amélioration à la fois algorithmiques, interactionnelles, matérielles et expérimentales :

- Sur le plan algorithmique, des techniques de compression telles que la distillation, l'élagage et la quantification devront être mises en œuvre afin de réduire la complexité de YOLOv7-Plus proposé et de l'architecture DINOv2+ConvLSTM, en vue d'un déploiement outdoor sur des plateformes embarquées plus performantes. Pour le système indoor, l'intégration d'une mémoire temporelle permettrait de stabiliser la classification des pièces, tandis que l'ajout d'un

magnétomètre et d'un baromètre améliorerait l'estimation d'orientation et la détection des changements de niveau.

- Sur le plan interactionnel, l'introduction de commandes vocales (pause, répétition, réglage du débit) et l'exploration de modalités complémentaires, telles que le retour vibrotactile, renforceraient la robustesse de l'assistance, notamment en environnement bruyant. La personnalisation des paramètres d'annonce contribuerait également à une meilleure adaptation aux besoins individuels.
- Sur le plan matériel, l'adoption de plateformes plus puissantes et de capteurs visuels à meilleure dynamique améliorerait les performances globales, tout en nécessitant une optimisation énergétique pour préserver l'autonomie.
- Enfin, une validation en conditions réelles d'usage, notamment à travers des études longitudinales impliquant des utilisateurs déficients visuels, constituera une étape déterminante pour confirmer la robustesse et l'acceptabilité du système dans des contextes quotidiens.

5.6. Conclusion

Ce chapitre a présenté l'évaluation des systèmes de navigation outdoor et indoor destinés à l'assistance des personnes déficientes visuelles, en analysant leurs performances respectives sur les plans perceptif, décisionnel et interactionnel.

Pour le système outdoor, les résultats ont confirmé la supériorité de YOLOv7-Plus sur les différentes bases de données testées, avec des gains significatifs en termes de précision et de mAP. Le module d'évitement à logique floue a démontré une adaptation cohérente aux configurations d'obstacles, tandis que la localisation par filtre de Kalman étendu avec fusion GPS/IMU a atteint des niveaux de précision supérieurs aux approches comparatives, validant la pertinence de la stratégie de fusion adoptée.

Concernant le système indoor, l'implémentation sur Raspberry Pi 4B a attesté de la faisabilité en temps réel. La classification des pièces, la détection d'objets et la localisation par fusion EKF-ultrasons ont présenté des performances compatibles avec un usage opérationnel. L'évaluation auprès d'utilisateurs déficients visuels a révélé une bonne acceptabilité et une charge cognitive modérée, confirmant l'intérêt fonctionnel du dispositif.

Malgré certaines limites, notamment l'absence de validation outdoor en conditions réelles et la nécessité d'une évaluation écologique approfondie en indoor, les résultats obtenus démontrent la viabilité technique et la cohérence globale des choix méthodologiques. Ils constituent une base solide pour l'évolution vers un système embarqué pleinement opérationnel pour la navigation assistée en environnements intérieurs et extérieurs.

Conclusion Générale

Conclusion Générale

Cette thèse a été consacrée à la conception d'un système intelligent d'assistance à la navigation destiné aux personnes déficientes visuelles, en adoptant une approche intégrée combinant la perception artificielle, la fusion multi-capteurs et la décision adaptative. L'objectif principal a consisté à proposer une architecture cohérente, capable de couvrir à la fois les environnements outdoor et indoor, tout en respectant les contraintes du temps réel et de l'embarqué.

Pour l'environnement outdoor, le système a été structuré autour de quatre sous-systèmes complémentaires : détection, classification, évitement et localisation. Le module de détection identifie les obstacles présents dans la scène à l'aide de modèles d'apprentissage profond. Le sous-système de classification distingue spécifiquement les obstacles statiques des obstacles dynamiques, information essentielle pour adapter la stratégie d'évitement. Le module décisionnel génère ensuite des trajectoires cohérentes en fonction de cette distinction et de la configuration spatiale observée. Enfin, la localisation repose sur une fusion GPS/IMU implémentée via un filtre de Kalman étendu, permettant une estimation continue de la position en présence d'incertitudes de capteurs.

Les performances des quatre sous-systèmes ont été évaluées séparément en environnement simulé. Les résultats obtenus démontrent la cohérence algorithmique de l'architecture proposée et la pertinence des choix méthodologiques. Toutefois, l'absence de validation en conditions réelles constitue une limite identifiée de ce travail et représente une étape essentielle pour la consolidation future du système.

En environnement indoor, l'approche a été adaptée aux contraintes spécifiques liées à l'absence de GNSS et à la forte proximité des obstacles. Le système combine une classification de pièces, une détection d'objets optimisée pour le temps réel et une localisation relative par fusion EKF-ultrasons. L'implémentation complète sur Raspberry Pi 4B a permis de valider la faisabilité opérationnelle sous contraintes computationnelles strictes. Contrairement au système outdoor, le système indoor a fait l'objet d'une évaluation expérimentale incluant des participants déficients visuels, permettant d'analyser à la fois les performances techniques et l'acceptabilité du dispositif.

Les retours utilisateurs ont mis en évidence l'importance du contrôle interactionnel et de la gestion du flux d'information. Ces observations orientent directement les perspectives d'amélioration, en particulier pour le système indoor. L'intégration de retours haptiques complémentaires au canal audio permettrait de diversifier les modalités sensorielles et d'améliorer la robustesse en environnement bruyant. L'ajout d'un bouton physique permettant de mettre le système en pause ou de l'arrêter renforcerait la maîtrise utilisateur et la sécurité perçue. Par ailleurs, l'optimisation du rythme de restitution des informations, en adaptant dynamiquement la longueur et la fréquence des messages à la vitesse de déplacement, contribuerait à réduire la surcharge cognitive tout en maintenant une assistance réactive.

Conclusion Générale

L'élargissement de la base de données indoor à des environnements tels que centres commerciaux, boutiques et bibliothèques constitue également une perspective stratégique. Une telle extension améliorerait la capacité de généralisation du système et favoriserait une navigation autonome dans des espaces complexes et fortement fréquentés.

Pour le système outdoor, les perspectives d'évolution concernent principalement l'amélioration algorithmique et la future validation réelle. L'intégration de versions plus récentes des architectures de détection, telles que YOLOv12 [138] ou YOLOv13 [139], pourrait accroître la précision tout en optimisant l'efficacité computationnelle. Sur le plan de la localisation, l'adoption d'un filtre de Kalman non parfumé, à dire Unscented Kalman Filter (UKF) [140] [141] ou Cubature Kalman Filter (CKF) [142] [143], permettrait une meilleure modélisation des non-linéarités et une estimation plus robuste en conditions dynamiques. Une validation en environnement réel contrôlé constituera une étape indispensable pour évaluer la performance globale en situation opérationnelle.

En définitive, cette thèse établit un cadre méthodologique structuré pour la conception de systèmes intelligents d'assistance à la navigation destinés aux personnes déficientes visuelles. Elle démontre la faisabilité d'une architecture modulaire intégrant détection, classification dynamique/statique, évitement et fusion probabiliste, tout en soulignant l'importance d'une validation progressive et centrée sur l'utilisateur. Les travaux réalisés constituent une base solide pour le développement futur de dispositifs plus robustes, multimodaux et pleinement opérationnels, susceptibles de contribuer durablement à l'autonomie et à l'inclusion.

Références

- [1] Muili, F., Okolie, A., Opere, C. A., Mbye, Y. F. N., & Ohia, S. E. (2025). Basic anatomy and physiology of the human eye. Dans *Research protocols for ophthalmic disease mechanisms and therapeutics: Glaucoma-ocular hypertension* (p. 1-8). Bentham Science Publishers.
- [2] Senoussi, M., & Dugué, L. (2019). La vision : un modèle d'étude de la cognition.
- [3] Kestens, C. (2019). Les déficiences visuelles. *Rev Med Brux*, 40, 278-284
- [4] Lee, S. Y., Gurnani, B., & Mesfin, F. B. (2024). Blindness. Dans *StatPearls*
- [5] Muhsin, Z. J., Qahwaji, R., Ghanchi, F., & Al-Tae, M. (2024). Review of substitutive assistive tools and technologies for people with visual impairments: recent advancements and prospects. *Journal on Multimodal User Interfaces*, 18(1), 135-156.
- [6] Real, S., & Araujo, A. (2019). Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors*, 19(15), 3404.
- [7] Awan, R., Fatima, G., & e Nayab, D. (2023). Investigation of blind spatial cognition and understanding of spaces to navigate without vision. *Review of Applied Management and Social Sciences*, 6(2), 351-360.
- [8] Congdon, N. G., Friedman, D. S., & Lietman, T. (2003). Important causes of visual impairment in the world today. *JAMA*, 290(15), 2057-2060.
- [9] Halligan, P. W., Fink, G. R., Marshall, J. C., & Vallar, G. (2003). Spatial cognition: evidence from visual neglect. *Trends in Cognitive Sciences*, 7(3), 125-133.
- [10] Hersh, M. (2020). Mental maps and the use of sensory information by blind and partially sighted people. *ACM Transactions on Accessible Computing (TACCESS)*, 13(2), 1-32.
- [11] Khan, S., Nazir, S., & Khan, H. U. (2021). Analysis of navigation assistants for blind and visually impaired people: A systematic review. *IEEE Access*, 9, 26712-26734.
- [12] Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4), 372-386.
- [13] Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. (2014). A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing Research*, 310, 60-68.
- [14] Kupers, R., & Ptito, M. (2014). Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neuroscience & Biobehavioral Reviews*, 41, 36-52.
- [15] Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, 134, 427-445.
- [16] Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, 380(6574), 526-528.
- [17] Shelton, A. L., & Yamamoto, N. (2021). Visual memory, spatial representation, and navigation. Dans *Current Issues in Memory* (pp. 27-63). Routledge.
- [18] Stroffregen, T. A., & Pittenger, J. B. (1995). Human echolocation as a basic form of perception and action. *Ecological Psychology*, 7(3), 181-216.
- [19] Thaler, L., & Goodale, M. A. (2016). Echolocation in humans: an overview. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(6), 382-393.
- [20] Vashist, P., Senjam, S. S., Gupta, V., Gupta, N., & Kumar, A. (2017). Definition of blindness under National Programme for Control of Blindness: Do we need to revise it? *Indian Journal of Ophthalmology*, 65(2), 92-96.
- [21] Jain, G., Teng, Y., Cho, D. H., Xing, Y., Aziz, M., & Smith, B. A. (2023). "I Want to Figure Things Out": Supporting exploration in navigation for people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1-28.
- [22] Strong, P. (2009). The history of the white cane. Tennessee Council for the Blind.
- [23] Pfaffenberger, C. J., Scott, J. P., Fuller, J. L., Ginsburg, B. E., & Biefelt, S. W. (1976). Guide dogs for the blind: Their selection, development, and training. Elsevier.
- [24] Kay, L. (1980). Air sonars with acoustical display of spatial information. In R. G. Busnel & J. F. Fish (Eds.), *Animal sonar systems* (pp. 769-816). Plenum Press.
- [25] Kay, L. (1973). The sonic glasses evaluated. *Journal of Visual Impairment & Blindness*, 67(1), 7-11.
- [26] Heyes, A. D. (1984). The sonic pathfinder: A new electronic travel aid. *Journal of Visual Impairment & Blindness*, 78(5), 200-202.

- [27] Smith, R. P., & Dailey, R. H. (1978). The Sonicguide: A valuable aid for blind children and infants. *Journal of Visual Impairment & Blindness*, 72(8), 317-319.
- [28] Joffe, E. (1987). Role of electronic travel aids: Field applications of the Russell Pathsounder. *Journal of Visual Impairment & Blindness*, 81(8), 389-390.
- [29] Kube, C. R. (1996). A minimal infrared obstacle detection scheme, *Practitioner*.
- [30] Besl, P. J. (1988). Active, optical range imaging sensors. *Machine Vision and Applications*, 1(2), 127-152.
- [31] Farcy, R., Leroux, R., Jucha, A., Damaschini, R., Grégoire, C., & Zogaghi, A. (2006, juillet). Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view. Communication présentée à la Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments (CVHI), Kufstein, Autriche.
- [32] Morrisette, D. L., Goodrich, G. L., & Hennessey, J. J. (1981). A follow-up study of the Mowat sensor's applications, frequency of use, and maintenance reliability. *Journal of Visual Impairment & Blindness*, 75(6), 244-247.
- [33] Benjamin, J. M., Jr., & Ali, N. A. (1974, mars). An improved laser cane for the blind. Dans *Quantitative imagery in the biomedical sciences II* (Vol. 40, pp. 101-104). SPIE.
- [34] Hofmann-Wellenhof, B., Lichtenegger, H., & Collins, J. (1997). Overview of GPS. Dans *Global Positioning System: Theory and practice* (pp. 11-26). Springer.
- [35] Golledge, R. G., Klatzky, R. L., Loomis, J. M., Speigle, J., & Tietz, J. (1998). A geographical information system for a GPS based personal guidance system. *International Journal of Geographical Information Science*, 12(7), 727-749.
- [36] Larsson, P., & Henriksson-Larsén, K. (2001). The use of dGPS and simultaneous metabolic measurements during orienteering. *Medicine and Science in Sports and Exercise*, 33(11), 1919-1924.
- [37] Holland, S., Morse, D. R., & Gedenryd, H. (2002). AudioGPS: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, 6(4), 253-259.
- [38] Garaj, V., Jirawimut, R., Ptasinski, P., Cecelja, F., & Balachandran, W. (2003). A system for remote sighted guidance of visually impaired pedestrians. *British Journal of Visual Impairment*, 21(2), 55-63.
- [39] Helal, A., Moore, S. E., & Ramachandran, B. (2001, octobre). Drishti: An integrated navigation system for visually impaired and disabled. Dans *Proceedings of the Fifth International Symposium on Wearable Computers* (pp. 149-156). IEEE.
- [40] Chen, X., Tremaine, M., Lutz, R., Chung, J. W., & Lacsina, P. (2006). AudioBrowser: A mobile browsable information access for the visually impaired. *Universal Access in the Information Society*, 5(1), 4-22.
- [41] Doobov, K. (2009). An overview of GPS systems and adaptations: Implications for the older user with vision impairment. *International Journal of Orientation & Mobility*, 2(1), 81-86.
- [42] Su, J., Rosenzweig, A., Goel, A., De Lara, E., & Truong, K. N. (2010, septembre). Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices. Dans *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 17-26). ACM.
- [43] Mulloni, A., Wagner, D., Barakonyi, I., & Schmalstieg, D. (2009). Indoor positioning and navigation with camera phones. *IEEE Pervasive Computing*, 8(2), 22-31.
- [44] Coughlan, J., & Manduchi, R. (2009). A mobile phone wayfinding system for visually impaired users. *Assistive Technology Research Series*, 25, 849-854.
- [45] Vlamincx, M., Jovanov, L., Van Hese, P., Goossens, B., Philips, W., & Pižurica, A. (2013, décembre). Obstacle detection for pedestrians with a visual impairment based on 3D imaging. Dans *2013 International Conference on 3D Imaging* (pp. 1-7). IEEE.
- [46] Bousbia-Salah, M., Bettayeb, M., & Larbi, A. (2011). A navigation aid for blind people. *Journal of Intelligent & Robotic Systems*, 64(3), 387-400.
- [47] Nguyen, Q. H., Vu, H., Tran, T. H., Van Hamme, D., Veelaert, P., Philips, W., & Nguyen, Q. H. (2014, septembre). A visual SLAM system on mobile robot supporting localization services to visually impaired people. Dans *European Conference on Computer Vision (ECCV)* (pp. 716-729). Springer.
- [48] Jafri, R., Ali, S. A., Arabnia, H. R., & Fatima, S. (2014). Computer vision-based object recognition for the visually impaired in an indoors environment: A survey. *The Visual Computer*, 30(11), 1197-1222.
- [49] Tapu, R., Mocanu, B., & Zaharia, T. (2017). DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance. *Sensors*, 17(11), 2473.

- [50] Endo, Y., Sato, K., Yamashita, A., & Matsubayashi, K. (2017, septembre). Indoor positioning and obstacle detection for visually impaired navigation system based on LSD-SLAM. Dans 2017 International Conference on Biometrics and Kansei Engineering (ICBAKE) (pp. 158-162). IEEE.
- [51] Zhang, H., & Ye, C. (2017). An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(9), 1592-1604.
- [52] Li, B., Munoz, J. P., Rong, X., Chen, Q., Xiao, J., Tian, Y., Arditi, A., & Yousuf, M. (2018). Vision-based mobile indoor assistive navigation aid for blind people. *IEEE Transactions on Mobile Computing*, 18(3), 702-714.
- [53] Wang, L., & Wong, A. (2019). Enabling computer vision driven assistive devices for the visually impaired via micro-architecture design exploration. *arXiv preprint*. <https://arxiv.org/abs/1905.07836>
- [54] Alhichri, H., Bazi, Y., Alajlan, N., & Bin Jdira, B. (2019). Helping the visually impaired see via image multi-labeling based on SqueezeNet CNN. *Applied Sciences*, 9(21), 4656.
- [55] Duh, P. J., Sung, Y. C., Chiang, L. Y. F., Chang, Y. J., & Chen, K. W. (2020). V-eye: A vision-based navigation system for the visually impaired. *IEEE Transactions on Multimedia*, 23, 1567-1580.
- [56] Vaishnav, D., Rama Rao, B., & Bade, D. (2020). Wearable assistance device for the visually impaired. Dans S. C. Satapathy, V. Bhateja, J. R. Mohanty, & S. K. Udgata (Éds.), *Advances in machine learning and computational intelligence: Proceedings of ICMLCI 2019* (pp. 667-676). Springer.
- [57] Afif, M., Said, Y., Pissaloux, E., & Atri, M. (2020, septembre). Recognizing signs and doors for indoor wayfinding for blind and visually impaired persons. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (pp. 1-4). IEEE.
- [58] Patel, I., Kulkarni, M., & Mehendale, N. (2024). Review of sensor-driven assistive device technologies for enhancing navigation for the visually impaired. *Multimedia Tools and Applications*, 83(17), 52171-52195.
- [59] Bamdad, M., Scaramuzza, D., & Darvishy, A. (2024). Slam for visually impaired people: A survey. *IEEE Access*, 12, 130165-130211.
- [60] Bougheloum, L., Bousbia-Salah, M., & Bettayeb, M. (2025, September). A Real-Time Spatiotemporal Classification System for Assisting Visually Impaired Individuals. In 2025 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1-8). IEEE.
- [61] Beingolea, J. R., Zea-Vargas, M. A., Huallpa, R., Vilca, X., Bolivar, R., & Rendulich, J. (2021). Assistive devices: Technology development for the visually impaired. *Designs*, 5(4), 75.
- [62] Delgrange, R. (2020). Améliorer la mobilité urbaine des personnes ayant un handicap cognitif en assistant la cognition spatiale : Analyse exploratoire de la chaîne du déplacement et évaluation expérimentale d'un prototype d'aide à la navigation, Thèse de doctorat, Université Paris Cité.
- [63] Islam, R. B., Akhter, S., Iqbal, F., Rahman, M. S. U., & Khan, R. (2023). Deep learning based object detection and surrounding environment description for visually impaired people. *Heliyon*, 9(6), e16924.
- [64] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., & Liu, H. (2021). Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(2), 109-127.
- [65] Chen, J., & Clarke, K. C. (2020). Indoor cartography. *Cartography and Geographic Information Science*, 47(2), 95-109.
- [66] Carlsson, G. (2020). Topological methods for data modelling. *Nature Reviews Physics*, 2(12), 697-708.
- [67] Nguyen, V. L., Yang, Y., & De Campos, C. (2023, Juillet). Probabilistic multi-dimensional classification. Dans *Uncertainty in Artificial Intelligence* (pp. 1522-1533). PMLR.
- [68] Ismail, A., Ahmad, S. A., Soh, A. C., Hassan, M. K., & Harith, H. H. (2020). MYNursingHome: A fully-labelled image dataset for indoor object classification. *Data in Brief*, 32, 106268.
- [69] Xia, H., Yao, C., Tan, Y., & Song, S. (2023). A dataset for the visually impaired walk on the road. *Displays*, 79, 102486.
- [70] Theodorou, P., & Meliones, A. (2023). Gaining insight for the design, development, deployment and distribution of assistive navigation systems for blind and visually impaired people through a detailed user requirements elicitation. *Universal Access in the Information Society*, 22(3), 841-867.
- [71] Sait, U., Ravishankar, V., Kumar, T., Bhaumik, R., KV, G. L., Bhalla, K., & Sanjay, K. S. (2020). Design and development of an assistive device for the visually impaired. *Procedia Computer Science*, 167, 2244-2252.

- [72] Yang, J., Wilde, A., Menzel, K., Sheikh, M. Z., & Kuznetsov, B. (2023, Septembre). Computer vision for construction progress monitoring: A real-time object detection approach. Dans L. M. Camarinha-Matos, A. Ortiz, & X. Boucher (Éds.), *Working Conference on Virtual Enterprises* (pp. 660-672). Springer Nature Switzerland.
- [73] Kuriakose, B., Shrestha, R., & Sandnes, F. E. (2022, Juin). LiDAR-based obstacle detection and distance estimation in navigation assistance for visually impaired. Dans M. Kurosu (Éd.), *International Conference on Human-Computer Interaction* (pp. 479-491). Springer International Publishing.
- [74] Malla, S., Sahu, P. K., Patnaik, S., & Biswal, A. K. (2023). Obstacle detection and assistance for visually impaired individuals using an IoT-enabled smart blind stick. *Revue d'Intelligence Artificielle*, 37(3), 571-578.
- [75] Song, Y., Li, Z., Li, G., Wang, B., Zhu, M., & Shi, P. (2023). Multi-sensory visual-auditory fusion of wearable navigation assistance for people with impaired vision. *IEEE Transactions on Automation Science and Engineering*, 21(3), 2487-2500.
- [76] Feghali, J. M., Feng, C., Majumdar, A., & Ochieng, W. Y. (2024). Comprehensive review: High-performance positioning systems for navigation and wayfinding for visually impaired people. *Sensors*, 24(21), 7020.
- [77] Duh, P. J., Sung, Y. C., Chiang, L. Y. F., Chang, Y. J., & Chen, K. W. (2020). V-eye: A vision-based navigation system for the visually impaired. *IEEE Transactions on Multimedia*, 23, 1567-1580.
- [78] Reyes Leiva, K. M., Jaén-Vargas, M., Codina, B., & Serrano Olmedo, J. J. (2021). Inertial measurement unit sensors in assistive technologies for visually impaired people, a review. *Sensors*, 21(14), 4767.
- [79] Liu, J., & Guo, G. (2021). Vehicle localization during GPS outages with extended Kalman filter and deep learning. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-10.
- [80] Ullah, I., Su, X., Zhang, X., & Choi, D. (2020). Simultaneous localization and mapping based on Kalman filter and extended Kalman filter. *Wireless Communications and Mobile Computing*, 2020(1), 2138643.
- [81] Xu, P., Kennedy, G. A., Zhao, F. Y., Zhang, W. J., & Van Schyndel, R. (2023). Wearable obstacle avoidance electronic travel aids for blind and visually impaired individuals: A systematic review. *IEEE Access*, 11, 66587-66613.
- [82] Mohd Romlay, M. R., Mohd Ibrahim, A., Toha, S. F., De Wilde, P., Venkat, I., & Ahmad, M. S. (2023). Obstacle avoidance for a robotic navigation aid using fuzzy logic controller-optimal reciprocal collision avoidance (FLC-ORCA). *Neural Computing and Applications*, 35(30), 22405-22429.
- [83] Liu, D., & Xu, L. (2025). Robot path planning and obstacle avoidance algorithm based on visual perception. *Neural Computing and Applications*, 37, 981-998.
- [84] Surougi, H. R., & McCann, J. A. (2023). Real-time optimisation-based path planning for visually impaired people in dynamic environments. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1839-1848). IEEE/CVF.
- [85] Sessner, J., Dellert, F., & Franke, J. (2022, Janvier). Multimodal feedback to support the navigation of visually impaired people. Dans **2022 IEEE/SICE International Symposium on System Integration (SII)* (pp. 196-201). IEEE.
- [86] GS, A. K., Pon, V. N., Rai, S., & Baskar, A. (2020). Vision system with 3D audio feedback to assist navigation for visually impaired. *Procedia Computer Science*, 167, 235-243.
- [87] Kleinberg, D., Yozevitch, R., Abekasis, I., Israel, Y., & Holdengreber, E. (2023). A haptic feedback system for spatial orientation in the visually impaired: A comprehensive approach. *IEEE Sensors Letters*, 7(9), 1-4.
- [88] Adebiyi, A., Sorrentino, P., Bohlool, S., Zhang, C., Arditti, M., Goodrich, G., & Weiland, J. D. (2017). Assessment of feedback modalities for wearable visual aids in blind mobility. *PLoS ONE*, 12(2), e0170531.
- [89] Bastola, A., Gluck, A., & Brinkley, J. (2023, Septembre). Feedback mechanism for blind and visually impaired: A review. Dans *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 67, No. 1, pp. 1748-1754). SAGE Publications.
- [90] Nebe, K., & Heimgärtner, R. (2024, Juin). Ergonomic principles in designing assistive systems. Dans M. Kurosu & A. Hashizume (Éds.), *International Conference on Human-Computer Interaction* (pp. 75-87). Springer Nature Switzerland.

- [91] Morris, L., Cramp, M., & Turton, A. (2022). User perspectives on the future of mobility assistive devices: Understanding users' assistive device experiences and needs. *Journal of Rehabilitation and Assistive Technologies Engineering*, 9, 20556683221114790.
- [92] Xie, X., Cheng, G., Wang, J., Li, K., Yao, X., & Han, J. (2024). Oriented R-CNN and beyond. *International Journal of Computer Vision*, 132(7), 2420-2442.
- [93] Rani, S., Ghai, D., & Kumar, S. (2022). Object detection and recognition using contour based edge detection and fast R-CNN. *Multimedia Tools and Applications*, 81(29), 42183-42207.
- [94] Vijayakumar, A., & Vairavasundaram, S. (2024). Yolo-based object detection models: A review and its applications. *Multimedia Tools and Applications*, 83(35), 83535-83574.
- [95] El-Saadawy, H., Tantawi, M., Shedeed, H. A., & Tolba, M. F. (2021, May). One-stage vs two-stage deep learning method for bone abnormality detection. In A. E. Hassanien, S. M. Elghamrawy, & I. Zelinka (Eds.), *Proceedings of the International Conference on Artificial Intelligence and Computer Vision* (pp. 122-132). Springer International Publishing.
- [96] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7464-7475). IEEE/CVF.
- [97] Bougheloum, L., Bousbia Salah, M., & Bettayeb, M. (2025). Real-time object detection for visually impaired people using an improved YOLOv7-Plus architecture. *Arabian Journal for Science and Engineering*, 1-18. Advance online publication.
- [98] Wang, Y., Sun, L., & Peng, D. (2022). A multihead convlstm for time series classification in ehealth industry 4.0. *Wireless Communications and Mobile Computing*, 2022(1), 8773900.
- [99] Mohd Noor, M. H., Tan, S. Y., & Ab Wahab, M. N. (2022). Deep temporal Conv-LSTM for activity recognition. *Neural Processing Letters*, 54, 4027-4049.
- [100] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., ... & Bojanowski, P. (2023). DINOv2: Learning robust visual features without supervision. *arXiv*.
- [101] Fu, Z. (2022). Vision transformer: Vit and its derivatives. *arXiv*. <https://arxiv.org/abs/2205.11239>
- [102] Han, D., Ye, T., Han, Y., Xia, Z., Pan, S., Wan, P., ... & Huang, G. (2024, September). Agent attention: On the integration of softmax and linear attention. In A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, & G. Varol (Eds.), *European Conference on Computer Vision* (pp. 124-140). Springer Nature Switzerland.
- [103] Oleiwi, B. K., Mahfuz, A., & Roth, H. (2021, January). Application of fuzzy logic for collision avoidance of mobile robots in dynamic-indoor environments. In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 131-136). IEEE.
- [104] Bougheloum, L., Salah, M. B., & Bettayeb, M. (2024, December). Fuzzy controller-assisted obstacle avoidance system for visual impairment. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1-6). IEEE.
- [105] Khodarahmi, M., & Maihami, V. (2023). A review on Kalman filter models. *Archives of Computational Methods in Engineering*, 30(1), 727-747.
- [106] Fang, H., Haile, M. A., & Wang, Y. (2021). Robust extended Kalman filtering for systems with measurement outliers. *IEEE Transactions on Control Systems Technology*, 30(2), 795-802.
- [107] Bougheloum, L., Bousbia-Salah, M., & Bettayeb, M. (2025, April). A visually impaired AI guided indoor system for room classification and depth-based localization. In B. A. Al-Matari (Ed.), *Proceedings of the International Conference on the Leadership and Management of Projects in the Digital Age* (pp. 293-307). Springer Nature Switzerland.
- [108] Upadhyaya, B. K., Pramanik, P. K. D., Roy, P., & Sen, R. (2024, January). Real-time obstacle detection using YOLOv8 on Raspberry Pi 4 for visually challenged people. In *International Conference on Smart Computing and Communication* (pp. 221-235). Singapore: Springer Nature Singapore.
- [109] Dong, K., Zhou, C., Ruan, Y., & Li, Y. (2020, December). MobileNetV2 model for image classification. In *2020 2nd International Conference on Information Technology and Computer Application (ITCA)* (pp. 476-480). IEEE.

- [110] Elngar, A. A., Arafa, M., Fathy, A., Moustafa, B., Mahmoud, O., Shaban, M., & Fawzy, N. (2021). Image classification based on CNN: A survey. *Journal of Cybersecurity and Information Management*, 6(1), 18-50.
- [111] Zainorzuli, S. M., Che Abdullah, S. A., Abidin, H. Z., & Ahmat Ruslan, F. (2022). Comparison study on convolution neural network (CNN) techniques for image classification. *Journal of Electrical and Electronic Systems Research (JEESR)*, 20, 11-17.
- [112] Afif, M., Ayachi, R., Said, Y., & Atri, M. (2020). Deep learning based application for indoor scene recognition. *Neural Processing Letters*, 51(3), 2827-2837.
- [113] Sohan, M., Sai Ram, T., & Rami Reddy, C. V. (2024). A review on YOLOv8 and its advancements. In *International Conference on Data Intelligence and Cognitive Informatics* (pp. 529-545). Springer.
- [114] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., ... & Ferrari, V. (2020). The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision*, 128(7), 1956-1981.
- [115] Sawicka-Gutaj, N., Gruszczyński, D., Guzik, P., Mostowska, A., & Walkowiak, J. (2022). Publication ethics of human studies in the light of the Declaration of Helsinki – A mini-review. *Journal of Medical Science*, 91(2), e700.
- [116] Qureshi, H. H., & Wong, D. H. T. (2020). Usability of user-centric mobile application design from visually impaired people's perspective. In M. Antona & C. Stephanidis (Eds.), *Universal Access in Human-Computer Interaction. Design Approaches and Supporting Technologies: 14th International Conference, UAHCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Proceedings* (pp. 311-322). Springer.
- [117] Bolton, M. L., Biltkoff, E., & Humphrey, L. (2023). The mathematical meaningfulness of the NASA Task Load Index: A level of measurement analysis. *IEEE Transactions on Human-Machine Systems*, 53(3), 590-599. <https://doi.org/10.1109/THMS.2023.3263482>
- [118] Heinold, E., Rosen, P. H., & Wischniewski, S. (2025). Usability questionnaire for robotic systems based on the ISO 9241-110. *IEEE Robotics and Automation Letters*, 10(3), 2231-2238.
- [119] Zhu, H., Wei, H., Li, B., Yuan, X., & Kehtarnavaz, N. (2020). A review of video object detection: Datasets, metrics and methods. *Applied Sciences*, 10(21), 7834.
- [120] Reddy, K. G., & Basha, S. S. (2025, March). Real time object identification: A study on COCO dataset. In *International Conference on Advanced Materials, Manufacturing and Sustainable Development (ICAMMSD 2024)* (pp. 860-870). Atlantis Press.
- [121] Al-refai, G., & Al-refai, M. (2020). Road object detection using Yolov3 and Kitti dataset. *International Journal of Advanced Computer Science and Applications*, 11(8).
- [122] Satarkar, H., Zagade, N., Gupta, S., & Pundlik, S. (2021). Comparative study between segmentation neural networks on pascal VOC dataset. *International Research Journal of Engineering and Technology*, 8(5), 3931-3935.
- [123] Oise, G. P., Unuigbokhai, N. B., Onwuzo, C. J., Nwabukei, O. C., Ejenarhome, P. O., Atake, O. M., & Bakare, S. K. (2025). YOLOv8-DeepSORT: A high-performance framework for real-time multi-object tracking with attention and adaptive optimization. *Journal of Science Research and Reviews*, 2(2), 92-100.
- [124] Kour, V., Kumar, S., Reddy, T. V., Poojary, K., Misra, R., & Singh, T. N. (2025). Exploring spatiotemporal relational learning with TimeSformer for identifying the severity of the road accidents. *IEEE Transactions on Computational Social Systems*. Advance online publication.
- [125] Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1-10.
- [126] Ab Wahab, M. N., Nazir, A., Ren, A. T. Z., Noor, M. H. M., Akbar, M. F., & Mohamed, A. S. A. (2021). Efficientnet-lite and hybrid CNN-KNN implementation for facial expression recognition on raspberry pi. *IEEE Access*, 9, 134065-134080.
- [127] Koonce, B. (2021). SqueezeNet. In *Convolutional neural networks with swift for tensorflow: Image recognition and dataset categorization* (pp. 73-85). Apress.
- [128] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., ... & Jain, M. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA realtime instance segmentation (Version v7.0) Computer software. Zenodo.

- [129] Alsultan, O. K. T., & Mohammad, M. T. (2023). A deep learning-based assistive system for the visually impaired using YOLO-V7. *Revue d'Intelligence Artificielle*, 37(4), 901-910.
- [130] Hidayatullah, P., Syakrani, N., Sholahuddin, M. R., Gelar, T., & Tubagus, R. (2025). Yolov8 to yolo11: A comprehensive architecture in-depth comparative review. *arXiv*. <https://arxiv.org/abs/2501.13400>
- [131] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82.
- [132] Kovacs, I. F., Karolyi, A. C., Stângaciu, C. S., Stângaciu, V., Nimară, S., & Curiac, D. I. (2025). An RFID-based indoor guiding system for visually impaired people. *Information*, 16(3), 220.
- [133] Feng, Y., Zhang, L., Wang, H., Chen, J., Liu, W., & Li, X. (2025). A smart cane-based indoor positioning system for visually impaired people. *IEEE Sensors Journal*, 25(7), 12085-12096.
- [134] Adapa, D., Shekhawat, V. S., Gautam, A., & Mohan, S. (2024). Visual assist system with enhanced localization for indoor navigation of visually impaired people. *IEEE Access*, 12, 162075-162093.
- [135] Tsai, C. H., Elyasi, F., Ren, P., & Manduchi, R. (2024). All the way there and back: Inertial-based, phone-in-pocket indoor wayfinding and backtracking apps for blind travelers. *ACM Transactions on Accessible Computing*, 17(4), Article 19, 1-35.
- [136] Zhang, K., Wang, Y., Shi, S., Li, J., Chen, M., & Liu, Y. (2024). Improved YOLOv5 algorithm combined with depth camera and embedded system for blind indoor visual assistance. *Scientific Reports*, 14, 23000.
- [137] Hsieh, Y. Z., Ku, X. L., & Lin, S. S. (2024). The development of assisted-visually impaired people robot in the indoor environment based on deep learning. *Multimedia Tools and Applications*, 83, 6555-6578.
- [138] Tian, Y., Ye, Q., & Doermann, D. (2025, February 18). YOLOv12: Attention-centric real-time object detectors. *arXiv*. <https://arxiv.org/abs/2502.12524>
- [139] Sanjalawe, Y., Makhadmeh, S. N., Almiani, M., & Al-E'mari, S. (2025). HyperFallNet: Human fall detection in real-time using YOLOv13 with hypergraph-enhanced correlation learning. *IEEE Access*, 13, 177111-177126.
- [140] Von See, T. B., Greinert, J., & Meurer, T. (2024). Detection and tracking of ocean layers using an AUV with UKF based extremum seeking control in the Baltic Sea. *Scientific Reports*, 14, Article 20629.
- [141] Feng, Y., Gao, M., Huai, W., Wang, W., Yang, N., & Sheng, L. (2025). UKF-based fault detection for nonlinear stochastic systems with strong noise and communication protocols. *ISA Transactions*, 167(Part A), 788-798.
- [142] Fu, H., Li, Z., Huang, W., Chengb, Y., & Zhang, T. (2023). A robust Cubature Kalman filter for nonlinear systems subject to randomly occurring measurement anomalies without a priori statistic. *ISA Transactions*, 139, 122-134.
- [143] Liu, D., Chen, X., & Cui, B. (2025). An improved Cubature Kalman filter for GNSS-denied and system-noise-varying INS/GNSS navigation. *Micromachines*, 16(10), 1116. <https://doi.org/10.3390/mi16101116>

Annexe A

Principes ergonomiques de la norme ISO 9241-110

La norme ISO 9241-110:2020, intitulée Ergonomie de l'interaction homme-système – Partie 110 : Principes d'interaction, constitue une référence internationale pour la conception et l'évaluation des systèmes interactifs. Elle définit un ensemble de principes ergonomiques permettant d'améliorer l'utilisabilité des interfaces utilisateur, indépendamment du domaine d'application ou de la technologie employée. Son objectif est de guider la conception d'interfaces offrant une interaction efficace, efficiente et satisfaisante pour les utilisateurs.

La norme est applicable à tous les systèmes interactifs et s'adresse aux analystes des besoins, concepteurs d'interfaces, développeurs, évaluateurs ainsi qu'aux organismes chargés de l'acquisition de systèmes interactifs. Elle constitue un cadre méthodologique permettant d'intégrer les exigences ergonomiques dès les premières phases de conception et tout au long du processus d'évaluation.

L'ISO 9241-110 repose sur sept principes fondamentaux d'interaction, qui doivent être pris en considération de manière conjointe lors de la conception ou de l'évaluation d'une interface utilisateur :

Adéquation aux tâches de l'utilisateur : le système doit permettre à l'utilisateur d'accomplir ses tâches de manière simple, efficace et avec un minimum d'effort.

Caractère autodescriptif : les informations nécessaires doivent être présentées clairement au moment opportun afin que l'utilisateur comprenne immédiatement l'état du système et les actions possibles.

Conformité aux attentes de l'utilisateur : le comportement du système doit être cohérent, prévisible et conforme aux conventions habituellement utilisées dans le contexte d'utilisation.

Facilité d'apprentissage : le système doit faciliter la découverte de ses fonctionnalités, limiter les besoins de formation et accompagner progressivement l'utilisateur dans son apprentissage.

Contrôle utilisateur : l'utilisateur doit conserver la maîtrise de l'interaction en pouvant interrompre, reprendre ou adapter les actions selon ses besoins.

Robustesse face aux erreurs d'utilisation : le système doit prévenir les erreurs, les détecter lorsqu'elles surviennent et offrir des mécanismes simples permettant leur correction.

Engagement de l'utilisateur : l'interface doit favoriser une expérience d'utilisation motivante, agréable et suscitant une interaction continue avec le système.

Ces principes ne sont pas indépendants les uns des autres et peuvent se compléter selon le contexte d'utilisation. Leur importance relative dépend des objectifs du système, des caractéristiques des utilisateurs, des tâches à réaliser ainsi que des contraintes liées à l'environnement d'utilisation. La norme souligne également que certains compromis peuvent être nécessaires afin d'optimiser l'utilisabilité globale du système.

Tableau A.1. Principes d'interaction et principales catégories de recommandations (ISO 9241-110)

Principe d'interaction	Catégories de recommandations
Adéquation aux tâches de l'utilisateur	a) détermination de l'adéquation du système interactif à une donnée b) optimisation de l'effort nécessaire pour accomplir la tâche c) valeurs par défaut permettant la prise en charge de la tâche
Caractère autodescriptif	a) présence et mise en évidence des informations b) indication claire de l'état du traitement
Conformité aux attentes de l'utilisateur	a) comportement et réponses appropriés du système b) cohérence (interne et externe) c) changements dans le contexte d'utilisation
Facilité d'apprentissage	a) découverte (des informations et des commandes que les utilisateurs recherchent) b) exploration (des informations et des commandes que les utilisateurs ont découvertes) c) conservation (des informations sur le système)
Contrôle utilisateur	a) interruption par l'utilisateur b) flexibilité c) individualisation
Robustesse face aux erreurs d'utilisation	a) prévention des erreurs d'utilisation b) tolérance aux erreurs d'utilisation c) correction des erreurs d'utilisation
Engagement de l'utilisateur	a) motivation de l'utilisateur b) fiabilité du système c) renforcement de l'implication de l'utilisateur vis-à-vis du système

Échelle d'évaluation utilisée

Dans le cadre de cette étude, l'évaluation ergonomique du prototype a été réalisée en s'appuyant sur les sept principes d'interaction définis par la norme ISO 9241-110. Les participants ont exprimé leur niveau d'accord avec chacune des affirmations proposées au moyen d'une échelle de Likert à cinq niveaux, permettant de mesurer leur perception de l'utilisabilité du système.

Tableau A.2. Échelle de Likert utilisée lors de l'évaluation

Valeur	Interprétation
1	Pas du tout d'accord
2	Plutôt en désaccord
3	Ni d'accord ni en désaccord
4	Plutôt d'accord
5	Tout à fait d'accord

Pour chaque affirmation, les participants sélectionnaient la réponse correspondant le mieux à leur perception après l'utilisation du système. Les réponses recueillies ont ensuite été exploitées afin de calculer les indicateurs statistiques (moyenne et écart-type) associés à chacun des sept principes ergonomiques, permettant ainsi d'évaluer le niveau global d'utilisabilité du prototype.