

Renforcement de la localisation piétonne par détection visuelle via des informations géométriques et spatiales contenues dans une carte

Mohamed Anis GHAOUI¹ Bastien VINCKE¹ Roger REYNAUD¹

¹SATIE—CNRS UMR 8029, Univ. Paris-Saclay, 91405 Orsay CEDEX, France

Résumé – Ce papier présente une technique pour renforcer un système proprioceptif de localisation piétonne en intérieur, assisté par une caméra, embarqué dans un téléphone. Un réseau de neurones YOLOv5 est entraîné pour détecter des amers présents dans les images. La principale contribution de ce papier est de permettre au système d’identifier les faux positifs en introduisant des contraintes géométriques et spatiales tirées de la carte du bâtiment. Ceci est accompli par un Filtre Particulaire avec gestion des particules en groupe validant ou non la détection dans un mode multi-hypothèses.

Abstract – This paper presents a technique to reinforce a proprioceptive system of indoor pedestrian location, assisted by a camera, embedded in a phone. A YOLOv5 neural network is trained to detect landmarks present in the images. The main contribution of this paper is to allow the system to identify false positives by introducing geometric and spatial constraints drawn from the building map. This is accomplished by a Particle Filter with management of particles in groups validating or not the detection in a multi-hypothesis mode.

1 Introduction

La navigation proprioceptive à base de données inertielles est une des modalités les plus utilisées en localisation pour piéton en intérieur [7]. Cette modalité à l’avantage d’être omniprésente dans les téléphones, matériellement peu coûteuse et non énergivore [11]. Mais, elle présente un bruit relativement important qui provoque la dérive dans le temps du cap du piéton et un cumul d’erreur sur la position du piéton [1]. Pour contrer cela, les systèmes de localisation proprioceptifs corrigent et recalent la position/cap via une observation extéroceptive de type radio-fréquence, lumière ou, dans notre cas, image [2, 10]. Un détecteur d’objet sur les images de la caméra du téléphone détecte la présence d’amer appartenant à des classes d’objets. Mais, le système de localisation est très sensible à un recalage visuel sur un faux positif [14]. Afin de palier ce phénomène, nous proposons d’utiliser la carte du bâtiment afin d’extraire des contraintes géométriques et spatiales permettant d’indiquer à quel point la détection est valide par rapport à l’information déjà présente.

2 Contexte du problème

Ce papier repose sur notre précédente publication [9] qui est résumée par la figure 1. Le piéton est équipé d’un téléphone ayant une acquisition des données d’accéléromètres et gyroscopes. La détection du *Pas* est composée de la détection de l’événement *Pas*, de la méthode de *Weinberg* pour l’estimation de sa longueur et du filtre de *Madgwick* pour le suivi de l’évolution du cap. Chaque événement *Pas* active une itération du filtre à particules pour estimer la nouvelle position et le cap de chaque particule.

Les particules ayant commis une infraction à la contrainte de mouvement en franchissant un mur sont éliminées (filtrage par murs). Si le nombre de particules après ce filtrage est inférieur à celui nécessaire pour le fonctionnement du PF, les particules manquantes sont régénérées afin de revenir à ce nombre.

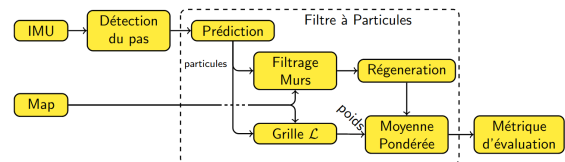


FIGURE 1 : Schéma bloc récapitulatif de [9].

En parallèle, les particules toujours valides tirent depuis une grille de vraisemblance leur poids. Cette grille est l’une des principales contributions de [9] où elle représente la vraisemblance qu’une particule prenne une position (x, y) , traduisant un comportement sur le déplacement dans l’environnement. La position de l’utilisateur est calculée via la moyenne pondérée à chaque pas.

Un comparatif de la présence de la grille et son absence est présenté dans [9]. Le système sans la grille est similaire à ceux présentés dans [13, 12]. Il est noté que la grille de vraisemblance n’a *a priori* pas d’impact sur la précision du système autour de 50 cm. La grille a par contre un impact évident sur l’incertitude autour de la position. Elle permet de resserrer les ellipsoïdes d’incertitude dans les couloirs ce qui limite leur élargissement dans les espaces plus ouverts.

Toutefois, cette grille provoque un effet indésirable lié à la configuration de l’espace en intérieur. Certaines parties de l’espace possèdent des murs asymétriques, des meubles arrangés non uniformément ou des obstacles. Ceci produit un biais dans la propagation des particules et les fait dévier vers l’espace plus ouvert, plus vite que sans la grille de vraisemblance. C’est pour cela que l’introduction d’une observation venant d’un capteur extéroceptif est nécessaire afin de : soit augmenter la précision du système ; soit réduire son incertitude.

3 Validation des détections

Nous nous plaçons dans un formalisme de Filtre Particulaire figure 2 où les blocs existant dans [9] sont en jaune, et les ajouts

en vert. Chaque particule peut procéder à la validation d'une détection et à partir de son *a priori* et de l'observation, calculer la vraisemblance *a posteriori* que la détection soit valide. Le processus de détection est réalisé avant chaque itération du PF. À chaque itération du PF, le processus de validation est exécuté. Les particules sont groupées suivant que la détection est supposée valide. Le fonctionnement multi-hypothèse du PF permet de différencier les traitements des différents groupes de particules.

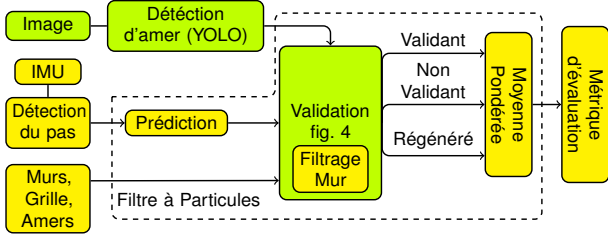


FIGURE 2 : Système avec blocs en jaune présent dans [9]. En vert, les ajouts.

3.1 Processus de détection

Afin de détecter des amers pendant la navigation, un détecteur d'objets dénommé YOLO est déployé. YOLOv5 [5] est un modèle basé sur des réseaux de neurones qui possède l'un des meilleurs compromis vitesse et performance. Ces réseaux ont montré de meilleurs résultats pour cet exercice [6] mais ils demandent des bases de données adaptées à l'environnement traité. L'extincteur est considéré comme un objet remarquable, non-intrusif car sa présence et sa disposition doivent répondre aux normes de sécurité anti-incendie. YOLO est entraîné sur le jeu de données COCO [8].

Ensuite, un apprentissage par transfert spécialisé YOLO sur une version de FireNet [3] que l'on a épuré. Seuls sont gardés les extincteurs encadré par une boîte et répondant à la norme européenne. L'apprentissage par transfert est nécessaire dû au fait que suite à cette étape d'épuration de FireNet, il ne reste que 365 images valides.

La cadence vidéo dans notre téléphone est de 25 fps via l'application MARSLogger [4]. YOLO effectue une détection sur chacune d'entre elle. Puis, la détection ayant la plus haute confiance entre 2 pas successifs ou dans la 0.5 s passée est sélectionnée. Ceci s'apparente à un maximum de vraisemblance des détections sur une fenêtre temporelle.

Lorsque YOLO infère la présence d'un extincteur dans l'image, il génère une BBox entourant un extincteur. Cette BBox est utilisée pour générer une pseudo-observation O_t qui sera utilisée pour un recalage visuel après validation :

$$O_t = [u, v, du, dv, x_g, y_g]^T \quad (1)$$

Où :

1. $[u, v] \in \mathbb{N}^2$ est la position du centre de l'amer dans l'image.
2. $[du, dv] \in \mathbb{N}^2$ est la largeur et la hauteur en pixels de l'objet détecté dans l'image.
3. $[x_g, y_g] \in \mathbb{R}^2$ la position de l'amer dans le plan du bâtiment.

3.2 Processus de validation

Nous proposons d'utiliser la carte du bâtiment afin d'extraire des contraintes géométriques et spatiales permettant d'indiquer

à quel point la détection en cours est valide par rapport à une détection visuelle O_t et à l'état de la particule P_t^i en cours de traitement (voir figure 4).

Plusieurs contraintes vont être étudiées et participer au calcul de la vraisemblance de la décision vrai ou faux positif de cette détection visuelle.

3.2.1 Validation par largeur de détection

À partir de la largeur de la détection du , on calcule à partir des bords gauche et droit de la boîte dans l'image, les angles, Ω_g et Ω_d via le modèle sténopé.

$$u_{g,d} = u \pm du/2 \rightarrow \Omega_{g,d} = -\arctan \frac{u_{g,d} - C_u}{F_u} \quad (2)$$

Où C_u et F_u sont les constantes du modèle sténopé pour les colonnes de l'image.

Depuis la position de chaque particule, deux rayons gauche et droit r_g et r_d sont lancés dans la carte et fournissent un point de collision K_g et K_d avec les murs par l'équation (3) en se basant sur la contrainte que les extincteurs sont toujours sur un support qui est vissé à la verticale du mur.

$$K_{g,d} = \begin{bmatrix} x_{g,d} \\ y_{g,d} \\ id_{g,d} \end{bmatrix} = \begin{bmatrix} R \cos(\theta^i + \Omega_{g,d}) \\ R \sin(\theta^i + \Omega_{g,d}) \\ \text{carte}(x_{g,d}, y_{g,d}) \end{bmatrix} \quad (3)$$

id est l'identifiant du mur auquel le point de collision appartient. Il est retourné par la fonction carte $(x_{g,d}, y_{g,d})$. R est un rayon de longueur infinie (dans la pratique, 100 m). Nous calculons la largeur L_e^i du possible amer placé contre le mur telle que :

$$L_e = \sqrt{(x_g - x_d)^2 + (y_g - y_d)^2} \quad (4)$$

La vraisemblance de la validité de la détection par sa largeur, \mathcal{L}_L , peut être calculée via l'équation (5), où L_g est la largeur physique d'un extincteur standard. Cette vraisemblance est présupposée gaussienne de centre L_g et d'écart-type σ_g .

$$\mathcal{L}_L^i \sim \mathcal{N}(|L_g - L_e|, \sigma_g) \quad (5)$$

3.2.2 Validation par la hauteur de détection

Pour calculer la vraisemblance de la validité de la détection par sa hauteur \mathcal{L}_H , il faut établir une méthode permettant de prédire la distance de la caméra jusqu'à l'amer détecté à partir de l'image. À première vue, cette tâche est impossible dû à la projection 3D-2D. Les hypothèses suivantes sont posées :

1. L'amer est de dimensions connues et constantes.
2. L'extincteur est toujours attaché au mur. Il apparaît toujours à la vertical sur l'image.

La fonction hauteur-distance f est construite par le protocole suivant : un extincteur est posé au sol contre un mur. Nous mesurons au télémètre laser les repères au sol des distances au mur. Ces valeurs formeront l'axe des abscisses dans le graphe d'étalonnage. Le téléphone est posé port USB sur le repère, en face de l'extincteur. L'application MARSLogger enregistre pendant 10 s avec les mêmes paramètres sténopés figés pour toutes les séquences d'acquisition. Toutes les acquisitions sont passées à YOLO pour produire les détections sur l'extincteur.

Seuls les vrais positifs sont conservés en aval de cette détection. La figure 3 montre deux fonctions d'étalonnage pour deux niveaux de confiance de détection.

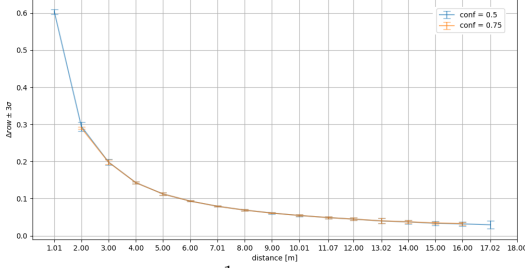


FIGURE 3 : Fonction f^{-1} d'étalonnage distance - hauteur.

Alors, le calcul de \mathcal{L}_H procède par :

1. Lancer un rayon dans l'angle $\theta^i + \hat{\Omega}^i$ i.e l'angle du cap dans lequel une particule prédit l'observation de l'amer potentiellement détecté.
2. Calculer la distance entre la particule et le point d'impact avec le mur, d_i .
3. Calculer \mathcal{L}_H qui est assimilée à une gaussienne de moyenne $f(du)$ et d'écart-type $\sigma_{g(du)}$.

$$\mathcal{L}_H^i \sim \mathcal{N}(|d_i - f(du)|, \sigma_{g(du)}) \quad (6)$$

3.3 Définitions des groupes

La validation multi-hypothèse d'une détection est schématisée dans la figure 4. Les processus de validations ont pour fonction de permettre de trier les particules en trois groupes, qui sont :

- Validant (V) : ce groupe de particules croient que la détection est correcte et utilise l'observation afin d'effectuer le recalage visuel.
- Non Validant (NV) : ce groupe ignore la détection et sa mise à jour fait décroître les poids des particules ou non.
- Invalides (I) : ce groupe a enfreint une contrainte de déplacement. Il est éliminé et doit être régénéré suivant [9].

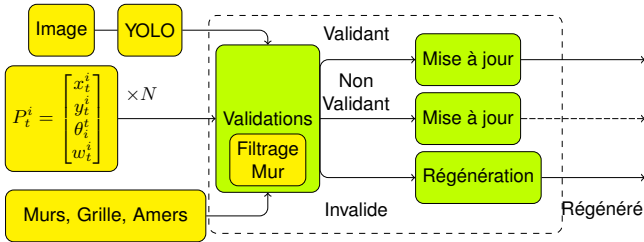


FIGURE 4 : Schéma bloc de l'intégration du processus de validation

Ces 3 groupes existent jusqu'à la prochaine détection où ils sont mélangés. En exploitant le poids d'une particule w_t^i et les vraisemblances introduites par équations (5) et (6), on peut calculer \mathcal{L}_d^i la vraisemblance de la détection d selon la particule i via équation (7).

$$\mathcal{L}_d^i \sim w_t^i \times \mathcal{L}_L \times \mathcal{L}_H \quad (7)$$

Enfin, une particule i est dans le groupe validant si $\mathcal{L}_d^i > \lambda$. Où, λ le seuil de validation. Sinon elle est dans le groupe non validant. La valeur de λ est $10^{-4} = (N \times 100)^{-1}$, soit 10 fois le seuil d'élimination d'une particule.

4 Résultats

La figure 5 est une visualisation des particules à un instant donné, chacune des 200 particules en bleu représentant une trajectoire possible dans le plan 2D du bâtiment. On peut constater que les particules se dispersent effectivement dans l'espace semi-ouvert mais avec une densité plus importante sur la droite. Le poids des particules à droite est aussi renforcé d'où la présence de pics magenta plus élevés sur celles-ci. Ces particules sont donc du groupe validant car elles ont été recalées visuellement sur un extincteur.

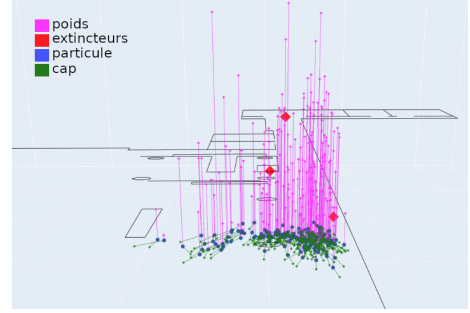


FIGURE 5 : Visualisation 3D des particules.

La figure 6 montre l'évolution du nuage à quelques instants où il y a eu une détection. Les particules en bleu sont du groupe validant. Les particules en rouge sont du groupe non validant. Le groupe invalide n'est pas affiché pour faciliter la lecture. Le groupe rouge est souvent présent contre le mur car il lui est impossible de valider la présence de l'amer parce que le lancé de rayon atterrit sur un obstacle. Quand une particule rouge n'est pas proche des murs, cela implique que son poids était faible.

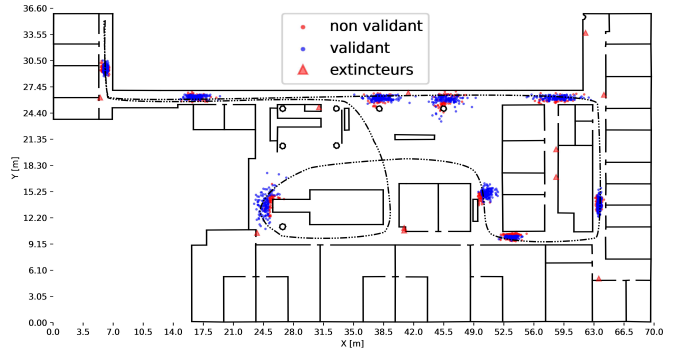


FIGURE 6 : Regroupement des particules à certains pas.

La figure 7 montre la répartition des 2 groupes bleu et rouge pour différent seuil de validation λ . λ peut être paramétré pour définir une confiance plus ou moins élevée dans les détections qu'il reçoit. La figure montre comment le seuil de validation de la vraisemblance de la détection fait évoluer les 2 groupes.

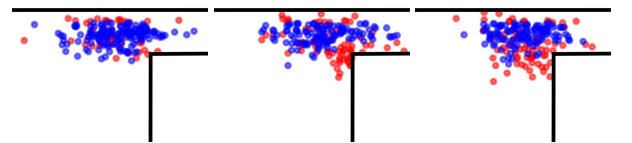


FIGURE 7 : Regroupement des particules à un pas donné pour λ , 2λ et 3λ .

Ceci fait de λ un paramètre de sensibilité et pourra donc servir à décrire un comportement plus évolué par l'avenir.

Les particules du groupe I sont éliminées et régénérées. Les particules du groupe V sont toujours visuellement recalées. Par contre, nous comparons 2 traitements sur NV. Ce groupe ignore la détection et sa mise à jour laisse le poids de ces particules inchangé ou l'annule. Dans le premier cas, elles continuent sans recalage visuel. Dans le second cas, le poids est annulé et les particules sont régénérées comme pour I.

Dans [9], 2 métriques, la distances euclidiennes (E) et de Mahalanobis (M), sont utilisées pour mesurer les performances. Elles sont formalisées par les équations (8) et (9).

$$E = \sum_{t=1}^K \frac{E_t}{K} \mid E_t = \sqrt{(l_t - \bar{P}_t)^2} \quad (8)$$

$$M = \sum_{t=1}^K \frac{M_t}{K} \mid M_t = (l_t - \bar{P}_t) \times \mathbf{S}_t^{-1} \times (l_t - \bar{P}_t)^T \quad (9)$$

Avec \bar{P}_t étant la position moyenne des particules $\mathbf{S} \in \mathbb{R}^{2 \times 2}$ étant la matrice de covariance des particules où chaque particule a deux composants, $P_t = (x_t, y_t)$:

$$\mathbf{S}_t = \mathbb{E} \left[\left(\mathbf{W}_t^T \mathbf{P}_t - \bar{P}_t \right) \times \left(\mathbf{W}_t^T \mathbf{P}_t - \bar{P}_t \right)^T \right] \quad (10)$$

La figure 8 montre une erreur euclidienne moyenne plus élevée pour le traitement annulé que pour le traitement inchangé. Les faux positifs présents dans l'expérience mettent en faute le groupe validant qui se recale dessus. L'information des particules qui n'avaient pas validé un faux positif est perdue selon le traitement annulé. Ce qui fait que le système de localisation est induit en erreur par leur absence.

Quant à la distance de Mahalanobis, le traitement annulé possède une plus grande distance car la distribution de ses particules possède une variance plus faible. Donc, le traitement annulé ne représente pas aussi bien l'espace des hypothèses que le traitement inchangé.

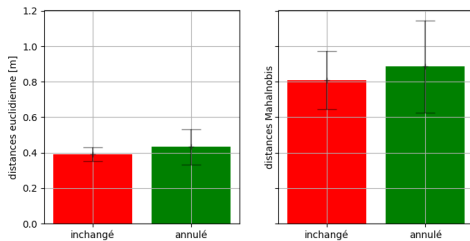


FIGURE 8 : Comparatif entre 2 traitements pour NV.

5 Conclusion et Perspectives

Nous avons montré dans ce papier comment le système de localisation peut corriger sa trajectoire et éviter la déviation présente dans [9] via le recalage visuel utilisant YOLO. Nous avons aussi présenté une approche pour définir 3 groupes de particules pour gérer les faux positifs et différencier les traitements. Cette approche permet d'utiliser le paramètre λ décrivant la sensibilité du système à ces détections.

Les perspectives incluent :

La carte peut être partagée par plusieurs piétons instrumentés car les extincteurs sont des amers bien définis. D'autres informations pourront être partagées au travers de la carte, comme la localisation des piétons. Ces éléments poussent nos

travaux vers la construction d'une modélisation de comportement par piéton.

L'analyse des faux positifs montre la présence d'humains parmi eux. La spécialisation au niveau du processus de détection d'objet a été trop forte. Il faut conserver différentes classes d'objets au moins extincteurs et humains pour diminuer le nombre de faux positif.

Références

- [1] A. Yassin et AL. : Recent advances in indoor localization : A survey on theoretical approaches and applications. *IEEE Communications Surveys Tutorials*, 19(2):1327–1346, 2017.
- [2] Beakcheol Jang et AL. : Survey of landmark-based indoor positioning technologies. *Information Fusion*, 89:166–188, 2023.
- [3] Jan Boehm et AL. : FireNet. <https://rdr.ucl.ac.uk/articles/dataset/FireNet/9137798>, 7 2019.
- [4] Jianzhu Huai et AL. : The mobile ar sensor logger for android and ios devices. *In 2019 IEEE SENSORS*, Oct 2019.
- [5] Jocher Glenn et AL. : yolov5. *Code repository www.github.com/ultralytics/yolov5*, 2020.
- [6] Krizhevsky Alex et AL. : Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [7] Kunthoth Jayakanth et AL. : Indoor positioning and way-finding systems : a survey. *Human-centric Computing and Information Sciences*, 10(1):1–41, 2020.
- [8] Lin Tsung-Yi et AL. : Microsoft coco : Common objects in context. *In European conference on computer vision*, pages 740–755. Springer, 2014.
- [9] Mohamed Anis Ghaoui et AL. : Human motion likelihood representation map-aided pdr particle filter. *IEEE Sensors Journal*, 2022.
- [10] Myriam Servières et AL. : Visual and visual-inertial slam : State of the art, classification, and experimental benchmarking. *Journal of Sensors*, 2021:1–26, 2021.
- [11] Nestor Michael Tiglaio et AL. : Smartphone-based indoor localization techniques : State-of-the-art and classification. *Measurement*, 179:109349, 2021.
- [12] Potortì Francesco et AL. : The ipin 2019 indoor localization competition—description and results. *IEEE Access*, 8:206674–206718, 2020.
- [13] V. Renaudin et AL. : Evaluating indoor positioning systems in a shopping mall : The lessons learned from the ipin 2018 competition. *IEEE Access*, 7:148594–148628, 2019.
- [14] Zhu Jiasong et AL. : Indoor topological localization using a visual landmark sequence. *Remote Sensing*, 11(1), 2019.