

Développement et évaluation des performances d'un filtre de Kalman sur une personne en mouvement

Nathan DAMETTE
Département informatique - Vision artificielle
Université de technologie de Belfort-Montbéliard (UTBM)
nathan.damette@utbm.fr

RÉSUMÉ

Ce document est un rapport de travail pratique visant le développement d'un filtre de Kalman pour le suivi d'une personne en mouvement, ainsi que l'évaluation des performances de ce filtre. Les mesures nécessaires au fonctionnement du filtre ont été effectuées en observant les mouvements d'une personne dans une vidéo à l'aide de l'outil de vision artificielle YOLO.

1. INTRODUCTION

Ce rapport explore une approche combinant le filtre de Kalman aux algorithmes YOLOv8 et YOLOv8-Poses pour détecter et suivre une personne en mouvement, même lorsqu'elle est temporairement cachée derrière des obstacles. L'objectif est de maintenir une trajectoire robuste dans des scénarios du monde réel. L'évaluation des performances offre un aperçu des avantages et des limitations de la méthodologie, contribuant ainsi à l'évolution des techniques de suivi dans des environnements dynamiques. L'approche vise à suivre précisément une personne dans des vidéos dynamiques, même lorsqu'elle se cache momentanément derrière des obstacles. L'objectif est de maintenir une trajectoire robuste malgré les variations de visibilité. Cette étude évalue la performance de l'approche dans des conditions réalistes, offrant ainsi des implications pratiques pour des domaines tels que la surveillance vidéo et la robotique.

2. REVUE DE LA LITTÉRATURE

Le filtre de Kalman, conceptualisé par Rudolf E. Kalman dans les années 1960, est un élément fondamental dans le domaine de l'estimation et du traitement du signal, et ce, malgré son ancienneté. Depuis son introduction, cet algorithme a été largement adopté et a trouvé des applications significatives dans des domaines variés tels que la navigation, la robotique et la vision par ordinateur.

Les travaux initiaux, notamment l'article fondateur de Kalman ("A New Approach to Linear Filtering and Prediction Problems", 1960)[1], ont établi les bases théoriques qui ont soutenu le développement continu de cet outil. L'extension du filtre de Kalman aux systèmes non linéaires, par le biais du filtre de Kalman étendu (EKF) présenté par Kalman et Bucy ("New Results in Linear Filtering and Prediction Theory", 1961)[2], a considérablement élargi son domaine d'application.

L'évaluation rigoureuse des performances du filtre de Kalman revêt une importance cruciale. Des approches robustes ont été proposées, comme celles présentées par Julier et Uhlmann dans leur article de 1997 ("A Non-Divergent Estimation Algorithm in the Presence of Unknown Correlations")[3].

Dans le contexte de la robotique, le filtre de Kalman a trouvé une application significative dans le cadre du SLAM, où il est utilisé pour estimer simultanément la trajectoire d'un robot mobile et la carte de son environnement. Des contributions notables, telles que l'article de Smith, Self et Cheeseman ("Estimating Uncertain Spatial Relationships in Robotics", 1987) et celui de Dissanayake et al. ("A Solution to the Simultaneous Localization and Map Building (SLAM) Problem", 2001[4]), ont démontré l'efficacité du filtre de Kalman dans ce contexte spécifique. En conclusion, le filtre de Kalman est encore aujourd'hui un outil essentiel et polyvalent, dont l'application continue d'évoluer. Son impact significatif dans des domaines variés en fait un sujet d'intérêt continu pour la recherche scientifique.

3. MÉTHODOLOGIE

3.1 Vidéo utilisée

La vidéo utilisée pour ce projet est une séquence capturée en 4K à 60 images par seconde à l'aide d'un smartphone. Elle a été enregistrée à l'intérieur d'une salle carrée d'environ 10 mètres sur 10 mètres, vue depuis un des coins de la pièce. La caméra, bien que fixe dans son orientation générale, peut subir de légères perturbations au fil du temps.

Au centre de la salle, divers obstacles sont présents, tels que des chaises, des tableaux et des poteaux, créant ainsi un environnement dynamique. Ces éléments ajoutent un niveau de complexité au suivi de la personne en mouvement, qui utilise des "drift skates". Cette personne réalise deux tours complets de la salle, naviguant habilement autour des obstacles.

La mobilité constante de la personne crée des situations où elle peut disparaître partiellement ou complètement derrière les obstacles, mettant ainsi à l'épreuve la robustesse du suivi. Les légères perturbations dans l'orientation de la caméra ajoutent également une dimension réaliste au défi du suivi dans un environnement intérieur. Cette étude évalue la performance de l'approche dans des conditions réalistes, offrant ainsi des implications pratiques pour des domaines tels que la surveillance vidéo et la robotique.

3.2 Filtre de Kalman

3.2.1 Fonctionnement Fondamental

Le filtre de Kalman repose sur un processus itératif qui combine les informations provenant de deux sources principales : les mesures actuelles et les prédictions basées sur l'état précédent du système. Le filtre équilibre de manière optimale ces deux sources d'information, tenant compte des incertitudes associées aux mesures et aux modèles dynamiques.

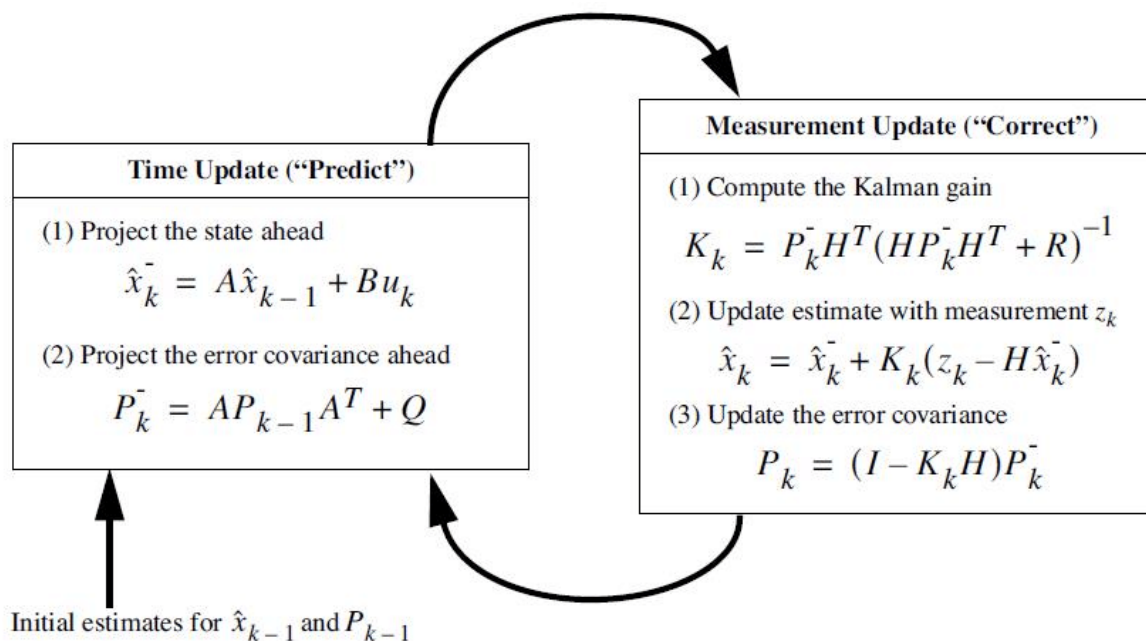
Prédiction de l'État

À chaque itération, le filtre de Kalman prédit l'état futur du système en utilisant les équations de transition dynamique. Cette prédiction est basée sur l'état estimé précédent et le modèle dynamique du système.

Mise à jour par Mesure : L'algorithme utilise ensuite les nouvelles mesures pour corriger la prédiction initiale. La correction est effectuée en fonction de l'incertitude associée aux mesures et à la prédiction, assurant ainsi une estimation plus précise de l'état actuel.

3.2.2 Équations Clés du Filtre de Kalman

Les équations fondamentales du filtre de Kalman sont définies comme suit :



3.2.3 Avantages et Limitations

Le filtre de Kalman est apprécié pour sa capacité à fournir des estimations précises, même en présence de bruit. Cependant, il suppose souvent des modèles linéaires et des distributions gaussiennes, ce qui peut limiter son efficacité dans des situations non linéaires ou fortement bruitées.

3.3 Acquisition des observations

3.3.1 Présentation de YOLOv8

YOLOv8 (You Only Look Once Version)[5] est une version évoluée de l'algorithme de détection d'objets YOLO développé par Joseph Redmon, Ali Farhadi et Santosh Divvala en 2016. YOLO se repose sur le principe fondamental de détection d'objets en une seule passe, offrant des améliorations significatives par rapport à ses prédécesseurs.

YOLO est un algorithme de détection d'objets en une seule étape. Contrairement aux méthodes en deux étapes, il prédit simultanément les boîtes englobantes et les classes d'objets dans une seule itération, rendant le processus de détection plus rapide.

YOLO utilise un réseau de neurones convolutifs pour extraire des caractéristiques significatives de l'image. Ces caractéristiques sont ensuite utilisées pour prédire les positions et les classes des objets.

L'image d'entrée est divisée en une grille, et chaque cellule de la grille est responsable de la prédiction d'objets. Chaque cellule génère un ensemble de boîtes englobantes et de probabilités pour différentes classes.

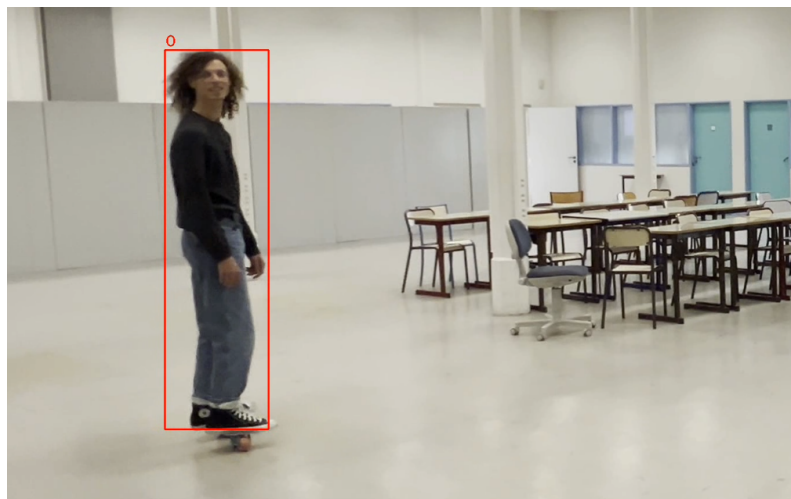


Figure 1 : YOLOv8 bounding box

En conclusion, YOLOv8, en tant qu'algorithme de détection d'objets rapide et précis, offre une combinaison puissante de précision et de vitesse, le positionnant comme une solution de choix pour une variété d'applications, notamment la détection d'objets en mouvement dans des vidéos dynamiques.

3.3.2 YOLOv8 version Poses

YOLO-pose[6] offre une avancée significative en détection d'objets en fournissant une solution intégrée pour la détection de personnes et l'estimation de leur pose 2D.

Contrairement aux méthodes traditionnelles qui se concentrent uniquement sur la bounding box, YOLO-pose va au-delà en identifiant les positions spécifiques des points clés liés au squelette de chaque personne détectée. Cette fonctionnalité permet une compréhension plus fine de la posture humaine, permettant de déterminer les coordonnées exactes de parties spécifiques du corps telles que les épaules, les coudes, les hanches et les genoux. Ainsi, en plus de localiser les individus dans une image, YOLO-pose offre une valeur ajoutée en permettant une analyse détaillée de la pose corporelle, ouvrant la voie à une gamme étendue d'applications telles que la surveillance de la santé, l'animation 3D, et l'interaction homme-machine basée sur la vision.

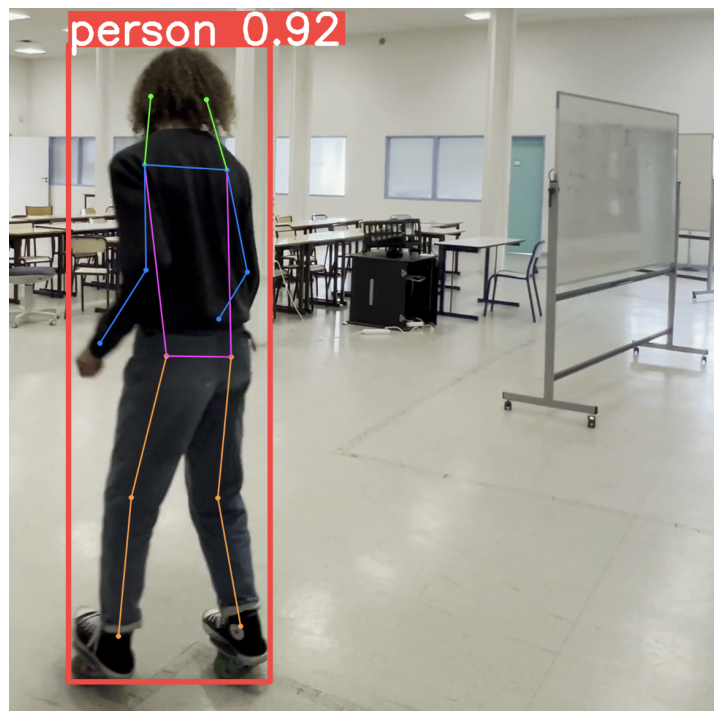


Figure 2 : YOLOv8-poses bounding box et keypoints

3.4 Méthode d'évaluation et de Mesure

3.4.1 Paramétrage et initialisation du filtre

Ce filtre de Kalman est adapté au suivi dynamique à 2 dimensions avec des paramètres soigneusement choisis pour une estimation précise de la position et de la vitesse d'un objet mobile.

```

kf = KalmanFilter(dim_x=4, dim_z=2)

# Define values for dt and the variances
dt = 1 / 60.0
sigma_pos_x = 1e-4
sigma_vel_x = 1e-2
sigma_pos_z = 1e-4
sigma_vel_z = 1e-2

sigma_pos_x_measure = 0.01
sigma_pos_z_measure = 0.1

# State transition matrix
kf.F = np.array([[1, dt, 0, 0],
                 [0, 1, 0, 0],
                 [0, 0, 1, dt],
                 [0, 0, 0, 1]])

# Measurement matrix
kf.H = np.array([[1, 0, 0, 0],
                 [0, 0, 1, 0]])

# Covariance matrices
kf.P *= np.array([[sigma_pos_x**2, 0, 0, 0],
                  [0, sigma_vel_x**2, 0, 0],
                  [0, 0, sigma_pos_z**2, 0],
                  [0, 0, 0, sigma_vel_z**2]])

# Measurement noise covariance matrix
kf.R *= np.array([[sigma_pos_x_measure**2, 0],
                  [0, sigma_pos_z_measure**2]])

# Process noise covariance matrix
kf.Q *= np.array([[sigma_pos_x**2, 0, 0, 0],
                  [0, sigma_vel_x**2, 0, 0],
                  [0, 0, sigma_pos_z**2, 0],
                  [0, 0, 0, sigma_vel_z**2]])

kf.x = np.array([initial_position[0], 0, initial_position[1], 0])
# print("kf.x", kf.x)

```

Figure 3 : Paramétrage et initialisation du filtre de Kalman

Le vecteur d'état, défini comme [position x, vitesse x, position z, vitesse z], englobe de manière concise la position et la vitesse selon les axes X et Z. Cette représentation complète vise à saisir de manière exhaustive la dynamique du mouvement.

La matrice de transition d'état (kf.F) orchestre la progression du vecteur d'état à chaque pas, prenant en compte la discrétisation du temps (dt). La matrice de mesure (kf.H) stipule le lien entre le vecteur d'état et l'espace des mesures.

Les matrices de covariance (kf.P, kf.R, kf.Q) sont des éléments cruciaux pour la fiabilité du filtre. La variance ajustée des positions et des vitesses sur les axes X et Z exprime la confiance accordée à ces composantes dans l'estimation. La faible variance des mesures de position comparée à celle des vitesses traduit la confiance dans la précision initiale des positions.

Concernant l'initialisation, les vitesses initiales sont fixées à zéro non pas en raison d'une décision stratégique, mais plutôt du constat que les premières frames ne sont pas souvent suffisamment précises. Cette approche repose sur l'espoir que le filtre convergera naturellement vers le modèle désiré à mesure que de nouvelles observations seront intégrées.

3.4.2 Constance des résultats

La première méthode d'évaluation repose sur la constance des résultats, visant à évaluer la cohérence entre les prédictions du filtre de Kalman et les observations réelles. Pour ce faire, les coordonnées prédites par le filtre de Kalman sont comparées à celles issues des observations, notamment les positions détectées par les algorithmes YOLOv8 et YOLOv8 poses.

En ce qui concerne les observations avec YOLO, les coordonnées sont obtenues en prenant les points déterminant la bounding box, ainsi que les coordonnées du point au centre de gravité de ce rectangle. Pour YOLO-poses, les coordonnées sont calculées en prenant les positions de deux points situés au niveau des hanches de la personne, et en déterminant les coordonnées du point situé au centre de ces deux points.

Étant donné que la personne en mouvement adopte des déplacements relativement fluides, on s'attend à ce que les courbes des positions prédites par le filtre de Kalman soient régulières et présentent une superposition avec les positions observées. Une concordance entre ces deux ensembles de données indiquerait la capacité du filtre de Kalman à suivre de manière précise et cohérente les mouvements de la personne dans la vidéo. Cette méthode d'évaluation vise ainsi à quantifier la fidélité du filtre de Kalman dans le suivi des trajectoires en mouvement, fournissant une mesure objective de sa performance dans des conditions dynamiques.

3.4.3 Evaluation par effacement des données

Dans le cadre de l'évaluation des performances du filtre de Kalman, une approche dynamique a été adoptée pour mettre à l'épreuve sa résilience et son adaptabilité dans des conditions de mouvement variées. Pour ce faire, nous avons utilisé la liste des coordonnées obtenues avec YOLOv8, représentant les positions détectées de la personne en mouvement. L'objectif était de simuler des situations où une partie des informations serait délibérément supprimée, reproduisant ainsi des scénarios où des obstacles, des changements de visibilité ou d'autres perturbations peuvent altérer la qualité des données d'observation.

La procédure de suppression d'information a été appliquée de manière sélective, volontairement omise pendant une seconde toutes les trois secondes, à partir de la deuxième seconde de la vidéo. Cette approche vise à évaluer la robustesse du filtre de Kalman dans divers états du mouvement, incluant des variations de vitesse, des trajectoires rectilignes ou en virage, ainsi que des trajectoires perpendiculaires ou parallèles à l'axe optique de la caméra.

Cette évaluation avec suppression d'information permettra d'analyser la capacité du filtre de Kalman à maintenir une prédiction précise et cohérente malgré des interruptions délibérées dans les observations, offrant ainsi des informations cruciales sur ses performances dans des conditions réalistes et dynamiques. Les résultats obtenus contribueront à une compréhension approfondie de la fiabilité du filtre dans des scénarios où la qualité des données d'observation peut varier significativement.

4. RÉSULTATS

4.1 Constance des résultats

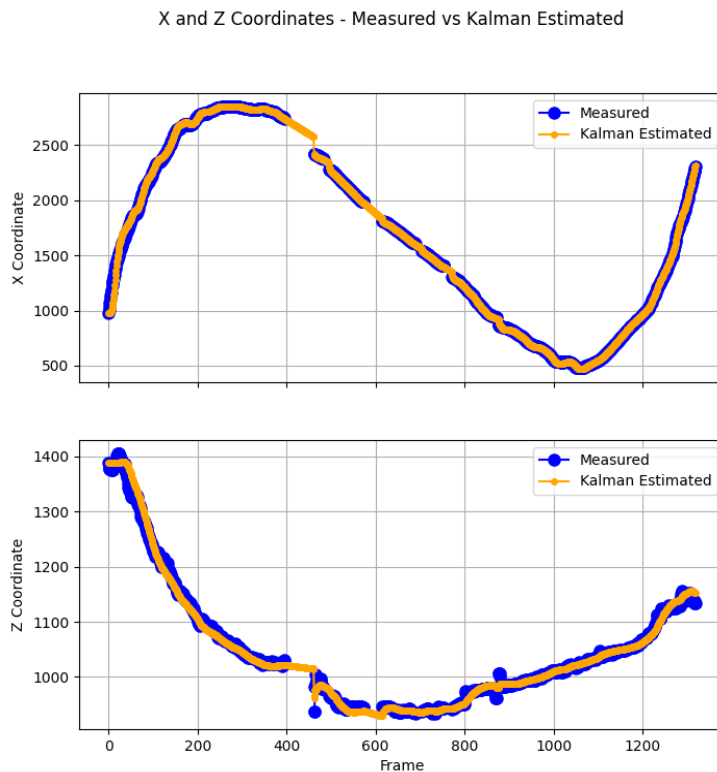


Figure 4 : Variation des coordonnées de Kalman et Yolov8

commence par rétrécir, créant ainsi un décalage du centre de gravité. Ce décalage est ensuite interprété comme une perte de vitesse par le filtre de Kalman, conduisant à une sous-évaluation de la variation de la position du sujet. Les fortes variations de l'observation en Z sont également dues à ce phénomène. Lorsque la personne observée passe derrière un obstacle tout en restant partiellement observable, la bounding box ne prend en compte que sa partie visible. Par exemple, lorsqu'elle passe devant une chaise, ses jambes ne sont plus prises en compte, entraînant une augmentation de la position du centre de gravité du rectangle sur l'axe Z.

Pour corriger ces erreurs liées à l'observation et à l'utilisation simple des bounding boxes, il a été nécessaire de trouver une solution pour utiliser les coordonnées d'un point déterminable présent sur la personne en mouvement. Les coordonnées approximatives du centre de gravité de la personne ont ainsi été obtenues à l'aide de YOLOv8-Poses, qui permet d'obtenir les coordonnées des hanches du sujet. Le centre de gravité approximatif est alors déterminé en calculant le centre de ces points. Comme le montre la Figure 5, les variations de la position dans la partie observable ont été significativement réduites grâce à cette méthode. De plus, en cas de perte d'observation, bien que celle-ci soit plus importante en raison de la moindre précision de l'observation à une plus grande distance par rapport à YOLO classique, on constate que l'écart entre l'estimation du filtre de Kalman et l'observation est largement inférieur à la méthode précédente, traduisant une meilleure précision.

Sur la Figure 4, les premiers résultats du filtre de Kalman avec YOLOv8 semblent corrects sur les deux axes lorsque l'observation est possible. Cependant, des variations importantes sur les deux axes apparaissent, ce qui est anormal compte tenu de la fluidité des mouvements de la personne suivie. En cas de perte d'observation, le filtre présente toujours un retard sur la position de la personne, retard qu'il rattrape lorsque l'observation redevient possible. Cette latence est attribuable à la méthode d'observation utilisée. Lorsque le sujet passe derrière un obstacle, la bounding box qui le caractérise

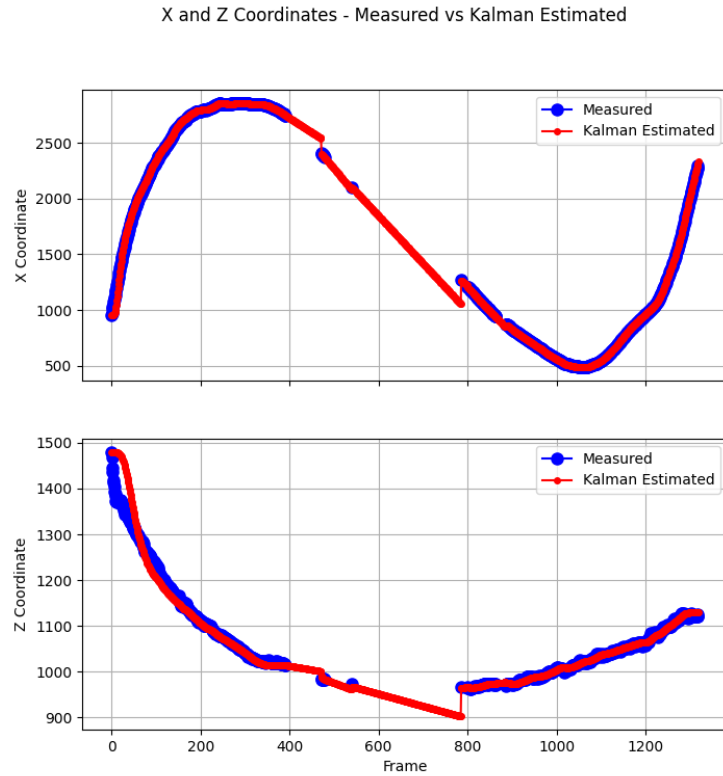


Figure 5: Variation des coordonnées avec YOLOv8-Poses

4.2 Evaluation avec effacement de données

La Figure 6 présente les résultats de l'évaluation des performances du filtre de Kalman lors de l'utilisation d'observations issues de YOLOv8-poses avec une suppression d'informations délibérée. Cette approche visait à simuler des conditions réalistes où une partie des données d'observation serait volontairement occultée. Les observations ainsi obtenues ont été partiellement effacées, avec une séquence d'une seconde exclue toutes les trois secondes, débutant à la deuxième seconde de la vidéo.

Les résultats mettent en lumière la robustesse du filtre de Kalman face à la perte d'observations dans des scénarios caractérisés par une variation constante de la position réelle. Cela correspond à des situations où la personne en mouvement maintient une vitesse constante ainsi qu'une trajectoire rectiligne et invariée. Dans de telles circonstances, le filtre parvient à prédire de manière précise malgré l'absence momentanée d'observations.

Cependant, il est important de noter que le filtre de Kalman montre des limitations lorsqu'une trajectoire change brusquement en l'absence d'observations. Ce comportement est intrinsèque à la nature du filtre, qui, en l'absence d'observations, s'appuie uniquement sur le modèle observé. Il estime la position du sujet comme s'il avait suivi une variation constante, ce qui peut conduire à des prédictions erronées lors de changements brusques de trajectoire.

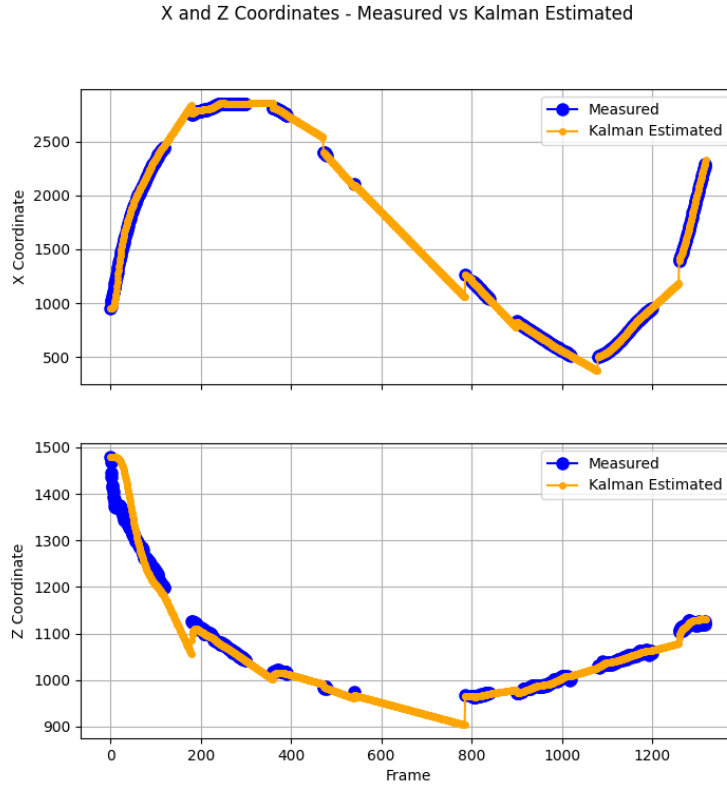


Figure 6: Variation des coordonnées avec suppression partielle des données

Cette observation souligne la nécessité d'envisager l'intégration d'une composante d'accélération dans le modèle du filtre de Kalman. Une telle amélioration pourrait permettre un suivi plus précis des objets non observables, caractérisés par un mouvement uniformément varié, c'est-à-dire une accélération constante. Cette considération pourrait enrichir la capacité du filtre à anticiper les mouvements dans des situations où les observations sont temporairement indisponibles.

4.3 Discussion des résultats

L'expérience menée souligne l'importance d'un modèle mathématique tel que le filtre de Kalman dans le contexte du suivi de personnes en mouvement, tout en mettant en lumière ses limites inhérentes. Plusieurs points significatifs émergent de cette évaluation, remettant en question la nature même des observations fournies.

Tout d'abord, l'observation n'est pas idéale, en particulier en ce qui concerne la composante Z qui est supposée rester relativement constante. Dans la réalité, la personne observée se déplace sur les axes X et Y. Il aurait été envisageable d'optimiser le filtre de Kalman en le restreignant au plan XY, en estimant la distance du sujet par rapport à la caméra pour déterminer ses composantes. Ceci aurait potentiellement réduit les variations de vitesse, étant donné que le sujet se déplace à une vitesse relativement constante le long des côtés d'un carré.

De plus, les résultats sont obtenus en comparant les sorties du filtre avec les observations fournies par YOLOv8. Cependant, l'exactitude de ces observations par rapport à la position réelle du sujet reste incertaine. Bien qu'une méthode manuelle de pointage sur la vidéo aurait pu fournir des positions réelles, elle ne garantit pas nécessairement une précision supérieure aux observations utilisées. La comparaison avec la réalité devient possible grâce aux graphiques, en prolongeant les points lorsqu'aucune observation n'est effectuée, étant donné que le mouvement suit des courbes sinusoïdales ou paraboliques.

Enfin, bien que YOLO-poses ait démontré des performances supérieures par rapport à YOLO simple, cette amélioration est associée à un coût computationnel considérablement plus élevé. Cette contrainte rend difficile l'utilisation de cette méthode en temps réel. Les limites de performance restent intrinsèquement liées aux contraintes réelles de coût, soulignant l'équilibre complexe entre précision et efficacité computationnelle dans le choix d'une méthode de suivi.

5. CONCLUSION

En résumé, cette étude a mis en lumière la robustesse du filtre de Kalman associé aux algorithmes YOLOv8 et YOLOv8-Poses pour le suivi de personnes en mouvement. Malgré des performances stables lors de mouvements constants, le filtre montre des limites face à des changements brusques de trajectoire sans observation. L'intégration de YOLOv8-Poses améliore significativement la précision. Cependant, le compromis entre performances et coûts computationnels, notamment avec YOLOv8-Poses, souligne les défis persistants dans le choix d'approches de suivi en temps réel. Ces résultats enrichissent notre compréhension des capacités du filtre de Kalman dans des scénarios réels, ouvrant la voie à des ajustements futurs pour une précision accrue.

6. RÉFÉRENCES

- [1] Kalman, R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82, 34-45.
<https://doi.org/10.1115/1.3662552>
- [2] Kalman, R. and Bucy, R. (1961) New Results in Linear Filtering and Prediction Theory. *ASME Transactions, Part D, Journal of Basics Engineering*, 83, 95-103.
<https://doi.org/10.1115/1.3658902>
- [3] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," *Proceedings of the 1997 American Control Conference* (Cat. No.97CH36041), Albuquerque, NM, USA, 1997, pp. 2369-2373 vol.4, doi:
<https://doi.org/10.1109/ACC.1997.609105>
- [4] M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," in *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229-241, June 2001.
<https://doi.org/10.1109/70.938381>
- [5] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics (Version 8.0.0) [Computer software].
<https://github.com/ultralytics/ultralytics>
- [6] Maji, D., Nagori, S., Mathew, M., & Poddar, D. (2022). YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss.
<https://doi.org/10.48550/arXiv.2204.06806>