

Comparaison de paramétrages du filtre de Kalman pour un suivi de personne dans une vidéo (avec occultation)

Nathan Rassié

Université de Technologie de Belfort Montbéliard, Belfort

Secteur Vision Artificielle



2 Décembre 2022

Résumé — Dans ce papier, il sera question d'effectuer le suivi d'une personne sur une trottinette électrique dans une vidéo, en étudiant différents paramétrages du filtre de Kalman classique. Nous utiliserons la Vidéo numéro 2 (fournie pour l'étude) sous trois framerate différents (60 / 30 / 10 FPS) et étudierons pour chaque déclinaison un modèle en position, en vitesse puis en accélération.

I. Introduction

Dans le cadre du cursus ingénieur en informatique à l'UTBM, le filtre prédictif de Kalman est étudié car il est un outil classique et très répandu depuis des dizaines d'années, que ce soit dans le monde de la vision par ordinateur ou même dans l'électronique pour les asservissements.

Dans le but d'étudier son comportement, nous allons chercher à effectuer le suivi d'une personne sur une trottinette électrique dans un échantillon vidéo fourni (Vidéo 2). Celle-ci est en 1080p et sera déclinée en 60, 30 et 10 images par seconde (FPS) pour l'étude. La vidéo contient une séquence lors de laquelle la personne à suivre est occultée par un poteau, ce qui permettra le filtre de Kalman à l'épreuve.

Ce papier s'articulera en trois axes : tout d'abord une discussion sur les choix d'implémentations (quelles technologies ont été utilisées, comment a été mise en place la détection ainsi que comment a été paramétré le filtre), puis une étude des mesures effectuées, pour finir sur une conclusion.

Une bibliographie se trouve à la fin de l'article. Le code source, les relevés de mesure, les graphiques, la vidéo originale ainsi que les images en taille originale se trouve en annexe à ce document, dans le même répertoire parent.

II. Implémentation

1. Choix des technologies

Pour réaliser cette étude, le choix du langage de programmation **Python** pour l'implémentation a été immédiat. En effet, lorsqu'il s'agit de faire un script court et rapide, ce langage est plus pertinent en terme de gain de temps qu'un langage bas-niveau, même si moins performant.

De plus, la librairie open-source de vision par ordinateur **OpenCV** a été utilisée dans ce cadre pour lire la vidéo, ainsi que pour son implémentation existante du filtre de Kalman.

Enfin, le logiciel *Numbers* d'Apple aura servi de tableur pour stocker les mesures et tracer les graphes, tandis que *Google Docs* aura permis la rédaction de ce papier en double colonnes.

2. Détection

La détection de la personne sur la vidéo permet, à chaque instant (donc à chaque image) d'obtenir sa nouvelle position (x,y) en pixels. Celle-ci nous permettra de mettre à jour le filtre de Kalman, pour que celui-ci corrige son modèle et effectue une prédiction plus précise au prochain instant.

La détection est donc très importante, cependant elle est aussi une problématique majeure en vision par ordinateur, et de nombreuses méthodes différentes existent pour obtenir de meilleurs résultats selon les contextes d'utilisation.

Or, le sujet de ce papier étant uniquement l'étude du filtre de Kalman, une perte de temps trop importante sur l'implémentation d'une excellente détection n'était pas envisageable.

De ce fait, une détection grossière par *binarisation d'image* puis par *ouverture* a été choisie. Celle-ci permet de supprimer les fenêtres apparentes lors de la *détection de contours*, au détriment du corps entier de la personne. Au final, seul le centre de la personne apparaît, mais cela a été jugé suffisant, puisque nous pouvons tout de même en récupérer le centre à chaque instant.



Figure 1 - Détection de la personne

Une autre approche aurait été de faire la différence entre la première image de la

vidéo (sans la personne) et toutes les autres où elle se déplace. Le point de vue étant statique, seule la personne se déplaçant aurait créé des différences d'intensités sur les pixels et aurait pu être facilement détecté.

Cependant, la mise au point permanente de l'appareil mobile utilisé pour effectuer la capture provoque de légers changements d'éclairages. De ce fait, la méthode par différence d'images pour retirer le fond échoue.

3. Paramétrage du filtre de Kalman

Le filtre Kalman d'OpenCV nécessite plusieurs paramètres pour fonctionner.

Tout d'abord, il faut définir sa matrice de mesure (H). Puisque nous ne connaissons que la position issue de la détection, nous n'avons que deux dimensions d'entrée, à savoir les coordonnées en x et en y du centre de la personne.

Ensuite, il faut établir la matrice de transition (A). C'est elle qui définira si l'on utilise un modèle en position, en vitesse ou en accélération (c'est à dire quelle composante on admet constante).

$$\begin{aligned}\ddot{x} &= a \\ \dot{x} &= a \cdot t + v_0 \\ x &= \frac{1}{2}a \cdot t^2 + v \cdot t + x_0\end{aligned}$$

Equation 1 - Modélisation de la position, de la vitesse et de l'accélération

Enfin, il faut définir les matrices de covariance de bruit sur la mesure et sur l'estimation (R et Q). Il s'agit, pour vulgariser, de définir le niveau de confiance que l'on en a en la mesure ainsi qu'en la prédiction du filtre.

Ces valeurs se déterminent majoritairement de façon empirique, de ce fait

elles ont été choisies arbitrairement pour cette étude en tenant compte des retours visuels des premiers essais.

La détection ayant été faite de façon grossière, nous choisissons une valeur de 500 pixels pour la matrice de bruit sur la mesure, pour tenir compte des grosses fluctuations en fin de vidéo. Pour la prédiction, étant donné que la trajectoire est simple avec peu d'obstructions, la valeur de 80 pixels sera retenue.

Autrement dit, nous considérerons pour ce test avoir plus confiance en la prédiction du filtre qu'en notre mesure. Ceci-dit, une étude entière pourrait être dédiée à l'évaluation des valeurs de bruits les plus optimales.

III. Mesures

1. Vidéo 2 - 60 FPS

Comme expliqué en introduction, la vidéo a été déclinée en trois fréquences d'images différentes. Regardons alors les résultats des différents modèles pour un extrait en 60 FPS, avec en rouge la mesure et vert la prédiction.



Figure 2 - Suivi en 60 FPS avec un modèle en Position constante



Figure 3 - Suivi en 60 FPS avec un modèle en Vitesse constante



Figure 4 - Suivi en 60 FPS avec un modèle en Accélération constante

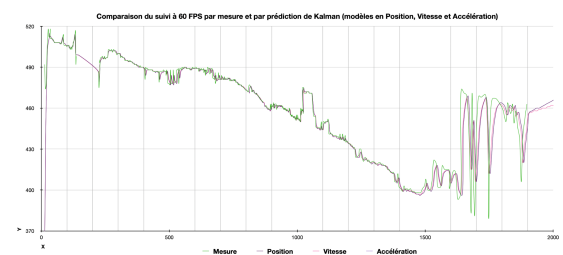


Figure 5 - Comparaison des coordonnées de la personne mesurées et déduites par Kalman

On remarque tout d'abord que le modèle en position est en retard sur la mesure (voir à droite la fin des tracés [Figure 2](#)). En effet, en considérant la position de la personne constante alors qu'elle ne l'est évidemment pas, le filtre se corrige à posteriori avec les nouvelles mesures et échoue à prédire les coordonnées future de la personne.

Pour les modèles en vitesse et accélération, le graphique [Figure 5](#) nous montre que les courbes sont confondues. En effet, la prédiction en vitesse constante est quasiment la même qu'en accélération.

Ceci s'explique car la vitesse de la personne est quasiment constante sur la vidéo,

ce qui mène donc à une accélération nulle d'après les équations

On notera également sur le graphe que lorsque la mesure échoue à cause de l'occultation provoquée par le poteau, le filtre prédit la même trajectoire peu importe le modèle.

2. Vidéo 2 - 30 FPS

Regardons maintenant les mesures sur l'échantillon à 30 FPS.



Figure 6 - Suivi en 30 FPS avec un modèle en Position constante



Figure 7 - Suivi en 30 FPS avec un modèle en Vitesse constante



Figure 8 - Suivi en 30 FPS avec un modèle en Accélération constante

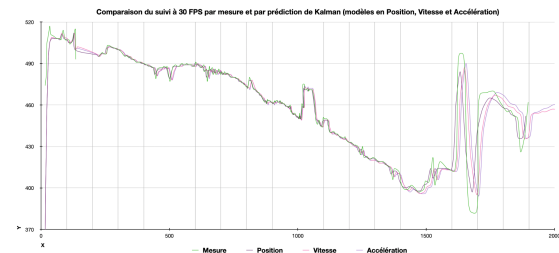


Figure 9 - Comparaison des coordonnées de la personne mesurées et déduites par Kalman

On remarque une nouvelle fois que la prédiction est en retard sur la mesure en modèle à position constante, pour les mêmes raisons qu'expliqué plus haut.

En revanche on distingue sur le graphe [Figure 9](#) une différence entre le modèle en vitesse et celui en accélération, ce-dernier étant légèrement plus en avance. Leur trajectoire reste cependant quasiment identique.

3. Vidéo 2 - 10 FPS

Même étude en 10 images par secondes (donc avec une vidéo très saccadée).



Figure 10 - Suivi en 10 FPS avec un modèle en Position constante



Figure 11 - Suivi en 10 FPS avec un modèle en Vitesse constante



Figure 12 - Suivi en 10 FPS avec un modèle en Accélération constante

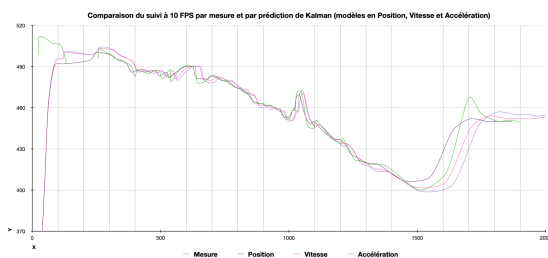


Figure 13 - Comparaison des coordonnées de la personne mesurées et déduites par Kalman

Les différences se font désormais beaucoup plus ressentir. En effet, le nombre réduit d'images provoque de plus grands écarts de la personne dans l'espace, et nous permet cette fois de bien discriminer les trois modèles.

Pour celui en position constante, toujours le même raisonnement qu'en 60 FPS.

En revanche, on distingue mieux sur le graphique [Figure 13](#) que, même s'il est plus en avance que le modèle en vitesse, le modèle en

accélération prédit une coordonnée en Y plus éloignée de la mesure.

En effet comme évoqué précédemment, la vitesse étant quasiment constante, il n'y pas réellement d'accélération de la personne sur la vidéo. Le modèle en accélération oscille donc en Y et s'éloigne plus que le modèle en vitesse de la mesure réelle.

IV. Conclusion

On réalise à l'issue de cette étude qu'un modèle en position constante tandis que l'objet à suivre se déplace est une aberration et fera échouer le filtre de Kalman à prédire ce que l'on souhaite.

On notera également qu'il est bien plus pertinent d'utiliser un modèle en vitesse puisque celle-ci est quasiment constante dans la vidéo. Ceci évitera des fluctuations trop fortes que pourrait causer le modèle en accélération.

Or, on observe également qu'avec un taux de rafraîchissement élevé, les modèles en vitesse et en accélération se confondent. Autrement dit, il y a assez d'images par secondes pour que la mesure corrige le filtre et l'empêche de trop dévier de la trajectoire réelle.

Ceci fonctionne probablement car, couplée aux 60 images par seconde, la personne se déplace lentement dans la vidéo. On pourrait cependant supposer que cela échoue même avec un haut taux de rafraîchissement si l'objet à suivre se déplacer puis vite ou de manière erratique.

Enfin, il est à noter que, lors du passage derrière le poteau, même si le filtre prédit une trajectoire suffisamment correcte, la vitesse n'est pas suffisante et son estimation doit être corrigée lorsque la mesure est

récupérée. Pour mieux comprendre cela, observer le tracé vert sur l'extrait "***Occultation.mov***" fourni avec ce papier.

Ceci pourrait s'expliquer soit par un mauvais paramétrage du filtre, soit parce que la personne passe derrière le poteau trop tôt dans la vidéo, donc avant que le filtre ait reçu assez de mesures pour se stabiliser.

V. Bibliographie

OpenCV Documentation - Kalman Filter

https://docs.opencv.org/3.4/dd/d6a/classcv_1_1KalmanFilter.html

Enseignement à l'UTBM - Cours de M.GECHTER sur le filtre de Kalman