



HAL
open science

Contrôle de caméra Pan-Tilt-Zoom réseau pour la poursuite d'une cible mobile

Pierrick Paillet, Romaric Audigier, Frédéric Lerasle, Quoc Cuong Pham

► **To cite this version:**

Pierrick Paillet, Romaric Audigier, Frédéric Lerasle, Quoc Cuong Pham. Contrôle de caméra Pan-Tilt-Zoom réseau pour la poursuite d'une cible mobile. Orasis, Congrès des jeunes chercheurs en vision par ordinateur, Jun 2013, Cluny, France. hal-00829309

HAL Id: hal-00829309

<https://hal.science/hal-00829309>

Submitted on 3 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contrôle de caméra Pan-Tilt-Zoom réseau pour la poursuite d'une cible mobile

P. Paillet¹

R. Audigier¹

F. Lerasle^{2,3}

Q.C. Pham¹

¹ CEA Nano-INNOV, Institut CARNOT CEA LIST, 91191 Gif-sur-Yvette Cedex

² CNRS, LAAS, 7 Avenue du Colonel Roche, F-31400 Toulouse

³ Université de Toulouse, UPS, LAAS, 118 route de Narbonne, F-31400 Toulouse

pierrick.paillet@cea.fr

Résumé

Les caméras Pan-Tilt-Zoom (PTZ) tendent à remplacer les versions fixes dans les situations de vidéo-surveillance car elles sont faciles à déployer, couvrent un plus large champ de vue et peuvent acquérir des images haute résolution de la cible grâce au zoom.

Néanmoins, les algorithmes existants couplant suivi de cible et automatisation des mouvements d'une PTZ ne sont pas robustes aux délais induits par les transmissions réseaux et les déplacements de la caméra. Nous proposons dans cet article une nouvelle stratégie de contrôle tenant compte de ces délais, basée sur une modélisation efficace de la chaîne de transmission des ordres à la PTZ, et sur une étape de prédiction des mouvements de la cible dans le plan du sol.

Mots Clefs

Suivi, caméra pan-tilt-zoom, délais, prédiction, stratégie de contrôle.

Abstract

Pan-Tilt-Zoom cameras (PTZ) seem to replace static ones in videosurveillance areas as they are easier to deploy, with a larger field of view and can take high resolution pictures of targets thanks to their zoom. However actual algorithms mixing tracking and camera control may not be robust to delays due to network transmission and PTZ motion. We propose here a new motion control strategy considering inherent delays, based on an efficient model of order transmission to PTZ and on a target movement prediction step.

Keywords

Tracking, pan-tilt-zoom camera, delay, prediction, law of control.

1 Introduction

Le suivi d'objet à l'aide de caméras fixes est une problématique centrale dans la communauté vision par ordinateur, motivée par de nombreuses applications dans le domaine de la vidéosurveillance. Aujourd'hui les caméras mobiles

et en particulier les caméras Pan-Tilt-Zoom (PTZ), i.e. capables de pivoter horizontalement et verticalement autour de leur centre et de zoomer sur leur cible, tendent à remplacer les versions fixes dans les grands espaces publics, tels que les halls de gare ou les grands magasins. Facilement accessibles dans le commerce, ces caméras sont faciles à déployer et contrôlables par réseau. De plus, elles peuvent couvrir un plus large champ de vue en se déplaçant et acquérir des images haute résolution de la cible avec leur zoom. Ainsi, ces caméras PTZ permettent de restreindre l'instrumentation d'environnement pour couvrir une scène à large échelle.

Les caméras PTZ réseau du commerce imposent cependant de fortes contraintes sur le traqueur et nécessitent une stratégie de perception-action adaptée pour maintenir la cible dans le champ de vue de la caméra. Cette loi de commande doit également tenir compte d'importants délais variables liés au fonctionnement des actionneurs et au temps de transmission réseau, pouvant induire une perte de la cible lors des déplacements. Des actions intuitives pour un opérateur humain, comme déplacer la PTZ ou zoomer pour suivre la cible, doivent être alors anticipées pour éviter la perte de la cible.

Cet article est structuré comme suit. La section 2 décrit brièvement les traqueurs PTZ existants et lois de commande associées. Ensuite, la section 3 détaille rapidement les contraintes inhérentes à la mise en oeuvre des PTZ commerciales et présente notre approche pour résoudre le suivi de cible mobile à l'aide d'une seule de ces caméras. Notre approche est composée d'une étape de détection de cible associée à un processus de filtrage spatio-temporel stochastique de type filtre particulière. Nous construisons également une stratégie de contrôle prenant en compte les divers temps de latence dans la boucle de perception-action de la caméra afin d'anticiper les commandes à appliquer aux actionneurs. Cette démarche dans le cadre des approches monoculaires est, à notre connaissance, peu étudiée dans la littérature. De plus, notre approche tient également compte de la modélisation des déplacements de

la PTZ commerciale pour rendre plus robuste le suivi et tire parti des délais pour optimiser les temps de calculs du traqueur. Enfin des évaluations et une étude comparative avec d'autres approches de l'état de l'art sont proposées dans la section 4.

2 Etat de l'art

De nombreuses approches existent pour tenir compte des contraintes liées aux caméras PTZ. La plus courante consiste à faire collaborer une caméra fixe et une (ou plusieurs) PTZ [6] de façon à assurer le suivi de la cible indépendamment de la PTZ, tandis que celle-ci est assignée à d'autres tâches comme l'acquisition d'images hautes résolutions du visage de la cible. Plusieurs PTZ peuvent également jouer alternativement le rôle de la caméra fixe pour permettre plus de flexibilité [8]. Cependant ces approches ne permettent pas de suivre une cible à l'aide d'une PTZ seule.

2.1 Suivi mono-caméra PTZ

La majorité des traqueurs actuels reposent sur la détection image par image et leur association dans le flux vidéo. Cependant celles-ci sont difficiles à obtenir avec une PTZ car on ne dispose pas de séquence off-line sur laquelle construire une modélisation de la scène comme souvent avec les caméras fixes.

La construction en amont d'un panorama [5, 9] qui recouvre l'ensemble des vues possibles de la PTZ alors qu'elle décrit un "tour de garde", permet de se replacer dans une situation où la soustraction de fond est applicable. Malheureusement on se heurte aux problématiques de mise à jour du modèle de scène car selon la durée et la taille de la zone surveillée, la PTZ peut mettre beaucoup de temps avant de percevoir une même scène. De plus, en situation de suivi de cible, le "tour de garde" doit être suspendu. Notre traqueur doit donc tirer parti des informations disponibles dynamiquement lors du suivi.

La mise en correspondance de points d'intérêt pour construire au fur et à mesure le modèle de scène [10] ou l'extraction de flot optique [2] permet de s'affranchir du "tour de garde". Cependant ces méthodes reposent sur des hypothèses de faibles changements d'apparences et de faibles déplacements entre images, ce qui est parfois difficile à maintenir selon le déplacement de la PTZ.

Inversement, les méthodes de détection de piétons à partir d'algorithmes d'apprentissage statistique [12] peuvent être utilisées dans toutes les circonstances, sans initialisation "off-line" préalable et sans conditions sur la fréquence d'acquisition de la PTZ ou ses déplacements. Elles nécessitent cependant davantage de calculs, mais ceci n'est pas limitant compte tenu des temps de latence importants des

PTZ commerciales, ce qui les rend plus pertinentes dans notre contexte.

2.2 Contrôle de la caméra

La stratégie de contrôle de la caméra et l'algorithme de suivi sont fortement dépendants l'un de l'autre. Le contrôle de la PTZ vise à garder la cible dans le champs de vue de la caméra ou à limiter la phase transitoire durant laquelle la cible n'est pas observée, soit en se déplaçant aussitôt que la position de la cible dans l'image est connue [3, 16], soit en trouvant un compromis entre nombre de déplacements (et donc accumulation de retard) et risque de sortie de la cible du champ de vue de la PTZ, généralement sous la forme d'une zone de "tolérance" autour du point visé par la caméra [2, 11].

Il est possible d'assembler un prototype de caméra directement à partir de ses composants, ce qui permet de contrôler et de modéliser séparément leur comportement et en particulier celui de ses servomoteurs actionnant tous les déplacements. Cette configuration permet, grâce à cette modélisation fine, d'utiliser l'ensemble des lois de commandes et en particulier de moduler la vitesse de déplacement de ce type de PTZ [1, 3].

Cependant avec des PTZ commerciales, il n'est pas possible d'avoir accès aux servomoteurs des caméras commerciales et de faire une modélisation fine de ses déplacements. Dans ce contexte, la commande en position est couramment utilisée [2, 13, 16] car elle permet d'assurer la position finale de la PTZ après l'exécution d'une commande. Cependant ces caméras ont souvent des temps de latence importants lors de l'exécution de la commande et de l'acquisition du flux vidéo, comme illustré dans la section 3.1, qui perturbent fortement les traqueurs utilisés. La prise en compte de ces délais dans le contrôle de la caméra, comme proposé dans [16], ainsi que dans le traqueur vise à améliorer la robustesse du suivi et constituera l'essence de notre approche.

Nous partageons plusieurs éléments de notre architecture avec [16] (voir figure 4), en particulier la prédiction de la position de la cible pour limiter l'influence des latences de la caméra. Cependant notre approche s'appuie sur une analyse plus fine du comportement de la PTZ et en tire davantage parti, comme nous allons le voir dans la partie suivante.

3 Notre approche

3.1 Caractérisation de la plateforme expérimentale

Nous utilisons en laboratoire une caméra AXIS 233D¹. Afin d'intégrer les délais de réaction de la PTZ dans

1. http://www.axis.com/fr/products/cam_233d/index.htm

notre stratégie, nous avons modélisé son comportement lorsqu'un ordre est donné de la manière suivante :

- Une phase d'attente, appelée également latence, durant laquelle l'ordre est transmis à travers le réseau puis de la partie logicielle de la PTZ vers ses actionneurs tandis qu'elle planifie son déplacement. La caméra elle-même reste statique.
- Une phase de déplacement effectif.

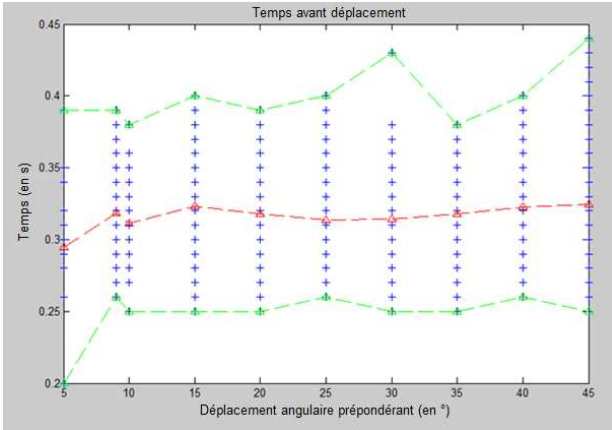


FIGURE 1 – Temps de latence mesuré sur la PTZ AXIS 233D en fonction de la valeur maximale du déplacement angulaire demandé. La courbe rouge représente la valeur moyenne et les courbes vertes les valeurs minimales et maximales.

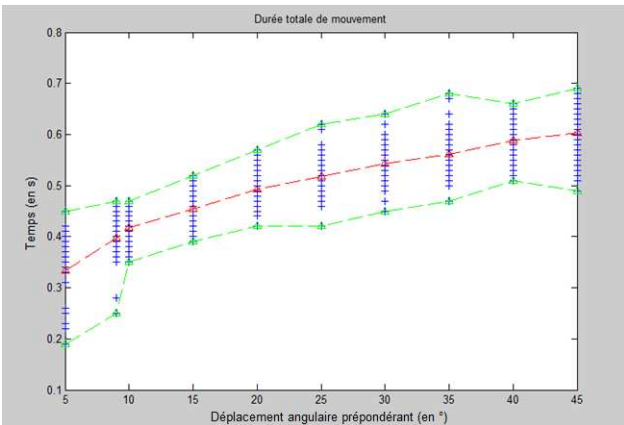


FIGURE 2 – Temps de déplacement mesuré sur la PTZ AXIS 233D en fonction la valeur maximale du déplacement angulaire demandé (en pan ou en tilt). La courbe rouge représente la valeur moyenne et les courbes vertes les valeurs minimales et maximales.

Nous avons adopté cette modélisation pour le comportement de plusieurs modèles de PTZ IP AXIS disponibles au laboratoire (AXIS 233D et Q6034²), même si les durées associées peuvent varier. Nous avons également constaté que les informations de positionnement en pan et

tilt lorsque la caméra se déplace ne sont plus forcément corrélées avec les images renvoyées. Ceci nous contraint à utiliser uniquement des informations image afin de ne pas être tributaires de l'implémentation interne de la caméra.

Pour notre caméra AXIS 233D, les figures 1 et 2 illustrent le temps de latence et le temps total d'exécution d'une commande mesuré en fonction de la valeur prépondérante $\max(\text{pan}, \text{tilt})$ de la commande. A l'instar de la littérature [12, 16], on observe des délais de déplacements compris entre 400 et 550 ms pour des commandes de 20° d'amplitude.

Cependant la figure 1 montre que pendant 300ms environ sur ce modèle de PTZ, la caméra est fixe et elle continue de nous envoyer des images nettes, donc exploitables.

Ceci nous amène donc à considérer quatre délais de natures différentes :

- τ_1 : Temps de transmission de l'image entre la PTZ et notre unité de traitement.
- τ_2 : Temps de traitement de notre traqueur.
- τ_3 : Temps de latence de la caméra, étudié figure 1.
- τ_4 : Temps d'exécution réel de la commande envoyée à la PTZ.

Le temps τ_2 est le seul dépendant de notre traqueur, τ_3 et τ_4 dépendent uniquement du modèle de PTZ utilisé et ont été étudiés en détail ci-dessus, tandis que τ_1 dépend de la configuration réseau sachant que, en général $\tau_1 \ll \tau_2, \tau_3$ et τ_4 .

Notre démarche vise donc à intégrer ces délais dans la stratégie de contrôle et à les exploiter pour améliorer les performances de notre traqueur, comme illustré figure 5.

3.2 Description de notre architecture

Notre architecture est illustrée figure 4; elle se compose d'un traqueur et de trois modules composant notre stratégie :

- Un module de prédiction qui va mettre à jour la position estimée par le traqueur et anticiper la position vers laquelle diriger la PTZ.
- Un module décisionnel, qui déclenche le déplacement de la caméra.
- Enfin un module d'interruption, qui évalue et corrige éventuellement la décision prise précédemment en exploitant le délai avant application effective de la commande.

Algorithme de suivi. Notre traqueur estime la position de la cible sur le plan du sol à partir d'une calibration géométrique [2] et d'une stratégie de filtrage particulière type "sampling-importance-resampling" (SIR).

Un détecteur de personnes basé apprentissage statistique [7] guide l'échantillonnage des particules du filtre dans

2. http://www.axis.com/fr/products/cam_q6034/

le plan du sol. Pour chaque particule, le modèle d'apparence combine un appariement de descripteurs de points d'intérêts SURF, pour tenir compte de la texture et d'histogrammes couleur HSV [14]. Ces derniers sont évalués sur un ellipsoïde 3D décomposé horizontalement en trois parties pour correspondre aux jambes, torse et tête de la cible. Cet ellipsoïde 3D est projeté dans le plan image de la PTZ comme illustré figure 3. Enfin le modèle est mis à jour par *boosting* pour tenir compte des changements d'apparence de la cible au cours du suivi.

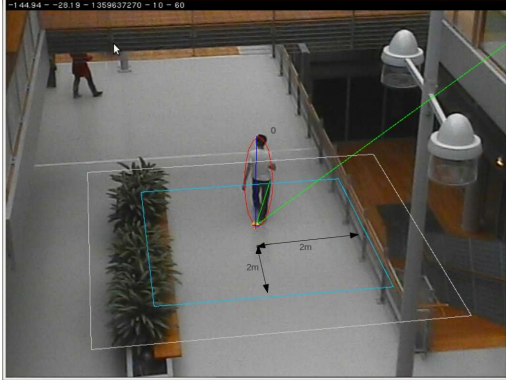


FIGURE 3 – Exemple d'image traitée par notre traqueur. La cible est modélisée par une ellipse rouge, la grille précalculée est indiquée en blanc. Lorsque la cible sort du rectangle bleu la PTZ se déplace.

Cependant la projection dans le plan image est couteuse en temps de calcul CPU, aussi nous discrétisons les positions que peut prendre la cible par une grille sur le plan du sol, avant d'effectuer la projection de l'ellipsoïde pour chaque élément de la grille. La projection ne dépend que des paramètres pan-tilt-zoom de la caméra, il est donc possible de conserver la même grille tant que la PTZ ne se déplace pas, accélérant l'évaluation des particules lors des itérations suivantes. De plus, comme la commande en position permet de connaître la position vers laquelle est dirigée la PTZ dès le moment où l'ordre est donné, nous pouvons effectuer le calcul de grille en parallèle pendant le déplacement (illustré figure 5).

Le raisonnement sur le plan du sol permet de tenir compte des informations 3D de la scène et modéliser plus facilement les mouvements de la cible. Enfin, dans la perspective d'un système de surveillance comprenant plusieurs capteurs, le référentiel lié au plan du sol est plus adapté pour échanger des informations sur la position des cibles.

Pour la suite nous adoptons les notations suivantes :

- X_k est la position de la cible au sol à l'instant t_k (vecteur d'état),
- Y_k est la position estimée par le traqueur de X_k (observation),

- $\hat{X}_{k+1|k}$ est la position prédite de X_{k+1} à t_k .

Module de prédiction. L'objectif de la prédiction est d'anticiper la future position de la cible X_{k+1} à l'instant $t_{k+1} = t_k + \sum_{i=1}^4 \tau_i$, connaissant sa trajectoire jusqu'à t_k . Il y a donc deux éléments à estimer :

- $\hat{\tau}_{\text{move}} = \hat{\tau}_3 + \hat{\tau}_4$: La prédiction s'effectue une fois la position Y_k connue, aussi les délais τ_1 et τ_2 sont connus. Mais les temps de déplacement τ_3 et τ_4 ne le sont pas.
- La transition $Y_k \rightarrow \hat{X}_{k+1|k}$: comment prédire le comportement de la cible en fonction des observations ?

L'analyse effectuée dans le paragraphe 3.1 permet d'avoir une bonne estimation de $\hat{\tau}_{\text{move}}$ en fonction du déplacement à réaliser. Par contre on ne dispose que de très peu d'informations a priori sur le comportement de la cible, qui doit alors être prédit à partir de la trajectoire déjà connue de la cible.

On peut faire l'hypothèse que la vitesse de la cible reste constante sur un petit intervalle de temps, comme le propose Varcheie et al. dans [16]. La moyenne glissante sur les N derniers vecteurs de mouvement permet alors d'avoir une première prédiction pour $\hat{t}_{k+1} = t_k + \tau_1 + \tau_2 + \hat{\tau}_3 + \hat{\tau}_4$:

$$\hat{X}_{k+1|k} = Y_k + \bar{v} \cdot \hat{\tau}_{\text{move}} \quad (1)$$

$$\text{où } \bar{v} = \frac{\sum_{i=1}^D \Delta X_i}{\sum_{i=1}^D \tau_1^i + \tau_2^i}$$

avec $D \leq N$ le nombre de vecteurs parmi notre historique restreint tels que l'angle entre deux vecteurs successifs soit inférieur à d° , $D \leq 3$, ΔX_i le vecteur de mouvement de la $i^{\text{ème}}$ itération précédente et τ_1^i et τ_2^i les délais associés.

La moyenne permet de lisser le bruit de mesure résultant de l'estimation de la position de la cible par le traqueur. Inversement la condition sur l'angle entre vecteurs permet de détecter un changement de dynamique et de s'y adapter. Cependant plus la trajectoire est complexe, plus le seuil d doit être faible pour que la prédiction soit réaliste. Or cela signifie que la moyenne est réalisée sur moins d'éléments et donc une plus grande sensibilité au bruit. De plus il sera difficile d'obtenir un compromis adapté à tous les contextes. C'est pourquoi nous avons remplacé cette prédiction avec moyenne glissante par un filtre de Kalman.

Le filtre de Kalman estime classiquement la probabilité conditionnelle $p(X_{k+1}|Y_{k+1})$ en deux étapes : une prédiction, permettant d'estimer $p(X_{k+1}|Y_k)$ à partir de $p(X_k|Y_k)$ et d'un modèle dynamique donné à priori ; et une mise à jour en intégrant la nouvelle observation Y_{k+1} , i.e. la sortie de notre traqueur à l'instant t_{k+1} . Si on pose l'hypothèse gaussienne que $\forall t, p(X_t|Y_t) = \mathcal{N}(\hat{X}_{t|t}, P_{t|t})$, on

retrouve les équations classiques :

$$\begin{aligned}
\hat{X}_{k+1|k} &= F_{k+1} \cdot \hat{X}_{k|k} && \text{(prédiction)} \\
P_{k+1|k} &= F_{k+1} \cdot P_{k|k} \cdot F_{k+1}^T + W && \text{(covariance prédite)} \\
\hat{Y}_{k+1} &= Y_{k+1} - H \cdot \hat{X}_{k+1|k} && \text{(innovation)} \\
S_{k+1} &= H \cdot P_{k+1|k} \cdot H^T + V && \text{(covariance de l'innovation)} \\
K_{k+1} &= P_{k+1|k} \cdot H^T \cdot S_{k+1}^{-1} && \text{(gain de Kalman optimal)}
\end{aligned} \tag{2}$$

Ce qui permet d'inférer l'état final et sa covariance associée :

$$\begin{aligned}
\hat{X}_{k+1|k+1} &= \hat{X}_{k+1|k} + K_{k+1} \cdot \hat{Y}_{k+1} && \text{(état mis à jour)} \\
P_{k+1|k+1} &= (Id - K_{k+1} \cdot H) \cdot P_{k+1|k} && \text{(covariance maj)}
\end{aligned}$$

Avec F_{k+1} la dynamique de notre modèle à l'instant t_k , W et V respectivement les matrices du bruit de dynamique et du bruit de mesure. Conformément à notre première hypothèse, nous avons choisi un filtre à vitesse constante entre chaque instant image. Cette hypothèse est également plus pertinente ici car l'intervalle de temps est plus court, *i.e.* une seule itération contre D itérations précédemment.

En conclusion, à chaque nouvelle image, la position Y_k obtenue par le traqueur permet de mettre à jour l'état courant $\hat{X}_{k|k}$ du filtre de Kalman à t_k . La prédiction à t_{k+1} , $\hat{X}_{k+1|k}$, permet ensuite d'anticiper le déplacement de la cible pour diriger la PTZ.

Module décisionnel. Idéalement, la loi de commande doit maintenir la cible au centre de l'image à chaque itération. Pour éviter de déplacer la PTZ sur des erreurs de suivi, Varcheie et al. [16] propose d'émettre une condition basée sur le score de similarité de l'apparence $App(t)$ de la cible associée au filtre à particules :

$$\text{tant que } App(t) \geq \alpha \cdot App_{ref} \text{ alors mouvement} \tag{3}$$

avec App_{ref} un score de référence et $\alpha = 85\%$ le seuil de décision.

Cette condition rend le système extrêmement sensible aux changements d'apparence. Selon la configuration, les cibles peuvent être éloignées de la caméra ou au contraire se rapprocher quasiment à la verticale, entraînant de fortes variations d'apparence de la cible, ce qui interdira à la PTZ de se déplacer (voir section 4). C'est pourquoi App_{ref} est actualisé par le score de similarité d'apparence de la cible après chaque déplacement.

Cependant en se déplaçant quasiment à chaque itération, cela revient à effectuer beaucoup de petits déplacements, réduisant considérablement les performances de l'algorithme de suivi (voir section 4). En relâchant partiellement la contrainte de centrage, *i.e.* en maintenant la cible dans un voisinage du centre, nous limitons le nombre de déplacements tout en étant plus robuste au bruit de mesure.

C'est pourquoi nous avons utilisé une stratégie, dite de "zone de tolérance" :

Soit X_{ref} la dernière position vers laquelle la PTZ a été dirigée, alors notre condition de déplacement s'exprime par :

$$\text{tant que } \|\hat{X}_{k+1|k} - X_{ref}\|_{\infty} < \epsilon \text{ alors pas d'ordre de déplacement} \tag{4}$$

Dans la pratique nous utilisons $\epsilon = 2m$.

Enfin notre module décisionnel est placé après le module de prédiction pour que ce soit la position prédite qui déclenche la condition de sortie de zone. Ainsi l'ordre de déplacement peut être émis en avance et tandis que la caméra se prépare, nous pouvons contrôler la qualité de notre prédiction avec le module d'interruption.

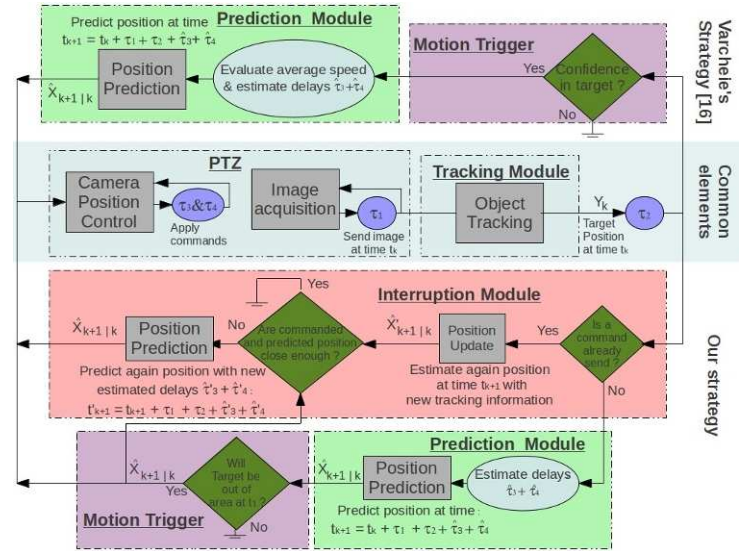


FIGURE 4 – Synoptique de notre architecture versus celle proposée dans [16].

Module d'interruption. Ce module est la principale innovation de cet article. En effet, à notre connaissance, rien dans la littérature n'est fait pour contrôler la position prédite vers laquelle la caméra va être dirigée une fois l'ordre donné. Or plus la durée d'estimation t_{k+1} est importante, plus l'erreur de prédiction commise est importante. Potentiellement, cela signifie diriger la PTZ vers une position où la cible sera hors de son champ de vue.

Lorsque l'ordre de se déplacer est donné, nous contrôlons l'erreur de prédiction en continuant à traiter les images tant que la PTZ ne s'est pas réellement déplacée (*i.e.* durant τ_3). Lors de ces itérations, nous effectuons de nouvelles prédictions $\hat{X}'_{k+1|k}$ au même temps t_{k+1} auquel la 1^{ère} prédiction a été faite. Celle-ci, réalisée avec de nouvelles données et sur un temps plus court, est plus fiable que la

précédente. (voir illustration figure 5).

Si la différence entre la position vers laquelle la PTZ a été dirigée $\hat{X}_{k+1|k}$ et celle nouvellement prédite $\hat{X}'_{k+1|k}$ est trop importante, on donne un contre-ordre pour rediriger la PTZ. Une nouvelle prédiction est réalisée avec les dernières observations et la PTZ est dirigée vers celle-ci. Notre temps de réaction est ainsi accru et diminue le risque de perdre la cible.

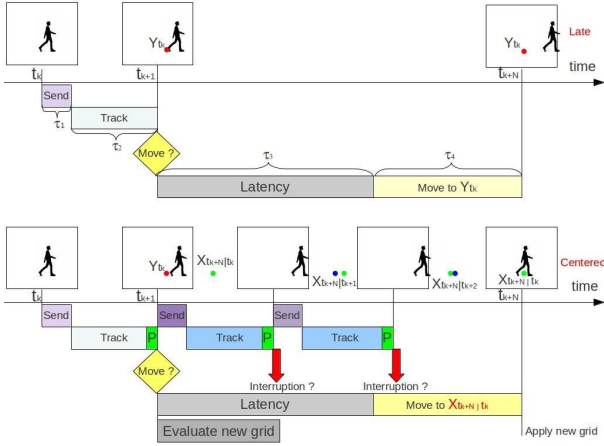


FIGURE 5 – Synoptique d’une boucle perception-action classique versus notre stratégie. En rouge la position obtenue par le traqueur, en vert la position prédite vers laquelle la PTZ est envoyée, en bleu les prédictions suivantes contrôlant la 1^{ère} prédiction.

4 Expérimentations et résultats

4.1 Protocole expérimental

Nous avons mené une série d’expériences afin de comparer deux stratégies de contrôle de la PTZ lors de mouvements standards :

Module	Stratégie n°1	Stratégie n°2
Prédiction	Vitesse Moy. (éq. 1)	Filtre Kalman (éq. 2)
Décision	Apparence (éq. 3)	Zone tolérance (éq. 4)
Interruption	Non	Oui

TABLE 1 – Stratégies de contrôle évaluées.

Afin de pouvoir comparer uniquement les deux stratégies, le traqueur décrit au paragraphe 3.2 est utilisé dans chacune des deux implémentations.

La première stratégie correspond à l’adaptation de la stratégie de contrôle de [16]. Cependant elle n’est pas employée dans le même cadre, en particulier le traqueur et le modèle d’apparence utilisés sont différents.

Nous avons évalué les deux stratégies sur différentes trajectoires au sol dont les caractéristiques sont données dans

le tableau 2. Les trois premiers scénarii sont des trajectoires simples, le quatrième scénario, illustré figure 7 est un mouvement plus complexe, qui inclut des demi-tours, des accélérations et des changements d’arrière-plan. Les deux derniers testent la robustesse du système face aux occultations liées à la présence d’autres personnes dans la scène, illustrés figure 6. Dans chaque cas, l’angle de vue sur la cible change fortement tandis qu’elle se rapproche ou s’éloigne de la PTZ.



FIGURE 6 – Résultats obtenus avec la stratégie n°2 pour les scénarii n°5 et n°6.

De part le comportement dynamique de la PTZ, on ne peut pas enregistrer et rejouer une même séquence pour évaluer nos stratégies. Il faut les évaluer via plusieurs réalisations de chaque scénario. Chaque scénario est donc rejoué cinq fois avec chaque stratégie. Pour reproduire autant que possible les mêmes conditions expérimentales, la trajectoire est marquée au sol et toutes les séquences ont été filmées à la suite. Chaque passage est enregistré avec la sortie du traqueur puis la vérité terrain est ensuite établie à la main sur les séquences obtenues.

4.2 Métriques

Pour caractériser les performances de notre système, nous avons privilégié les quatre métriques ci-dessous :

- *Pertinence* (Pert.) : ratio entre le nombre d’images correctement traitées et le nombre total d’images de la séquence. Par correctement traitée on entend que le recouvrement entre la boîte-cible obtenue par le traqueur et la vérité terrain 2D est supérieur à 50%.
- *Précision* (Préc.) : moyenne des ratios de recouvrement entre la boîte-cible et la vérité terrain 2D, sur le nombre d’images correctement traitées.
- *Focalisation* (Foc.) : moyenne sur les images contenant la cible du ratio entre la distance $f(\delta)$ et la diagonale de l’image. Avec δ la distance euclidienne entre le centre de la vérité terrain 2D et le centre de l’image et f :

$$\begin{cases} f(\delta) = 0 & \text{si } \delta < 25\% \text{ de la diagonale de l'image} \\ f(\delta) = \delta & \text{sinon} \end{cases}$$
- *Fragmentation* (Frag) : ratio entre le nombre d’images où la cible est hors de l’image, i.e. où plus de 50% de la vérité terrain est hors champ et le nombre total d’images.

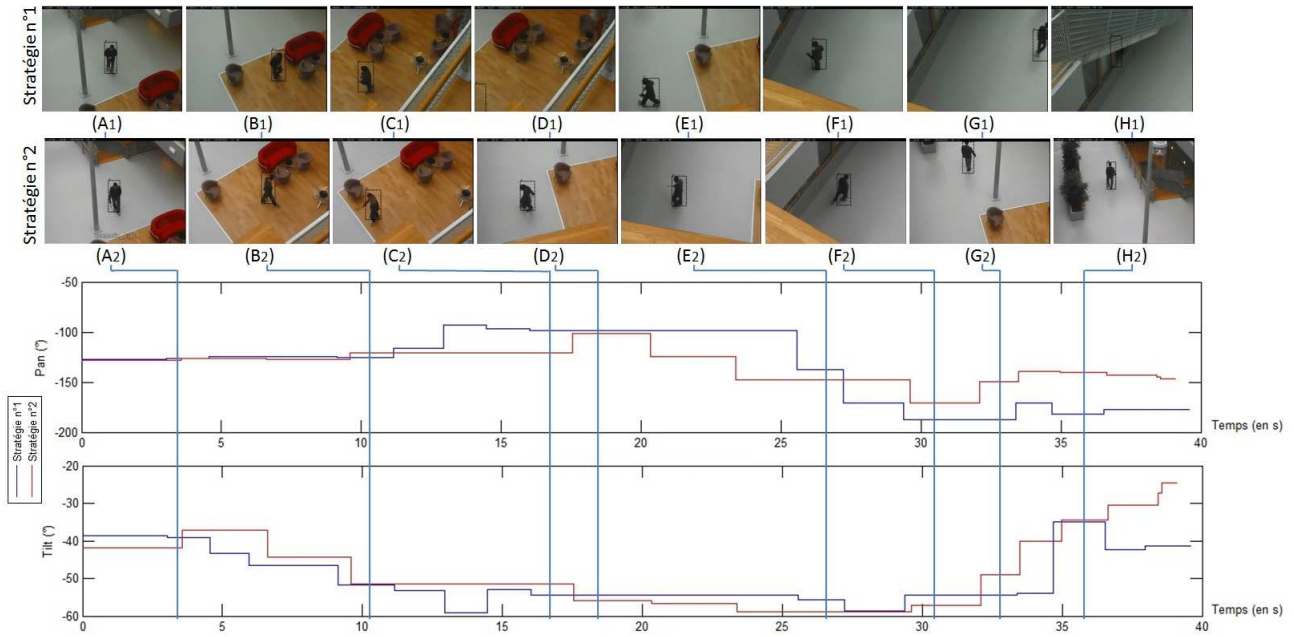


FIGURE 7 – Comparaison des résultats obtenus avec les deux stratégies pour le scénario n°4.

La pertinence et la précision sont issues des métriques de CLEARMOT [4] et rendent compte des performances du suivi de cible. Le traqueur étant identique pour les deux stratégies, on mesure ainsi l'influence du contrôle de la PTZ sur le traqueur. La focalisation et la fragmentation reflètent les performances du contrôle de la PTZ. La fonction de focalisation f permet de comparer les deux stratégies sans pénaliser la stratégie basée sur la zone de tolérance.

4.3 Résultats et discussion

Les résultats des expériences sont illustrés sur le tableau 2. La stratégie n°2 améliore les performances du traqueur en précision et pertinence dans tous les scénarii testés, i.e. elle lui fournit de meilleures conditions. D'autre part, on voit avec les scores de focalisation que les deux stratégies conservent de la même manière la cible dans la région centrale de l'image.

Par contre on constate que le nombre d'images traitées par seconde (fps) varie entre les scénarii et entre les deux stratégies. La différence entre scénarii s'explique par le nombre de déplacements nécessaires pour maintenir la cible dans le champ de vue. Par contre, le coût CPU des deux stratégies est relativement similaire car l'essentiel des calculs est effectué dans le traqueur, aussi l'écart entre stratégie provient majoritairement du nombre de déplacements effectués et de l'utilisation par la stratégie n°2 des images obtenues pendant le temps de latence.

La figure 7 illustre le déroulement des mouvements de la PTZ lors du scénario n°4 avec les deux stratégies. Sur les 20 premières secondes, on constate bien que le nombre de déplacements de la stratégie n°1 est plus important que pour la stratégie n°2, mais ceux-ci sont de faible amplitude. En particulier avec la stratégie n°2, la position de la PTZ est inchangée entre les figures 7.B₂ et 7.C₂.

Au delà des 20 premières secondes, la stratégie n°1 éprouve des difficultés pour remplir sa condition de déplacement basée sur l'apparence de la cible (éq. 3) à cause de l'angle de vue de la PTZ et du fond, capturé partiellement dans les histogrammes couleurs. Ces interférences empêchent le déclenchement du mouvement alors que la cible sort du champ de vue (figures 7.D₁ et 7.G₁), et conduisent finalement à la perte de la cible (figure 7.H₁).

Le même phénomène intervient lorsque la cible est occultée : des éléments des distracteurs vont venir perturber le modèle d'apparence de la cible. Lors du scénario n°5 (illustré figures 6.A₁ à 6.D₁) les deux personnes sont de couleurs proches, le traqueur peut dériver sur le distracteur mais le modèle d'apparence reste suffisamment proche de celui du dernier déplacement pour autoriser le déplacement, expliquant les bons résultats de la stratégie n°1. Dans le scénario n°6 les couleurs parasites sont différentes de celles de la cible (figure 6.B₂), modifiant le modèle d'apparence pendant l'occultation et empêchant la PTZ de la suivre après l'occultation (figure 6.C₂).

Enfin, la stratégie de déplacement par zone et l'interruption de commande permettent de mieux réagir face aux mouvements imprévisibles de la cible, comme le montrent les

Scénario n°	Type de mouvement	Nb personnes	Occultation	Durée	Pert.	Préc.	Foc.	Frag.	Fps	Echec
1	stratégie n°1 stratégie n°2	linéaire		24.24s	75%	71%	99%	0%	2.71	0
					85%	73%	100%	0%	5.59	0
2	stratégie n°1 stratégie n°2	chgt. brusque		25.27s	73%	71%	92%	6%	4.82	0
					88%	71%	99%	0%	6.05	0
3	stratégie n°1 stratégie n°2	circulaire		18.60s	70%	68%	92%	0%	2.5	0
					87%	72%	100%	0%	5.14	0
4	stratégie n°1 stratégie n°2	linéaire, chgt. brusque accélération, circulaire		39.17s	59%	65%	92%	20%	2.77	4
					72%	68%	97%	3%	4.63	1
5	stratégie n°1 stratégie n°2	linéaire, chgt. brusque	>40% sur ≈3s	23.73s	87%	70%	100%	0%	4.46	0
					93%	71%	98%	0%	5.80	0
6	stratégie n°1 stratégie n°2	linéaire	>75% sur ≈2s	16.67s	63%	65%	99%	18%	5.06	5
					87%	71%	98%	0%	6.51	0

TABLE 2 – Résultats moyens des deux stratégies pour chaque scénario.

meilleurs scores de pertinence et les taux de fragmentation plus faibles lors des scénarii n°2, 4 et 5.

5 Conclusion et perspectives

Notre approche dans cet article est structurée autour de la prise en compte des délais inhérents aux caméras PTZ. Les temps de calculs de notre traqueur sont optimisés en exploitant le temps de déplacement tandis que notre stratégie de contrôle repose sur trois améliorations eu égard à la littérature : une prise de décision basée sur la prédiction de la cible via un filtre de Kalman, une stratégie visant à restreindre le nombre d'ordres envoyées à la caméra et une gestion des erreurs commises lors de la prédiction.

Nous nous sommes inspirés des travaux de Varcheie et al.[16] qui, à notre connaissance, sont les seuls à prendre en compte les latences dans leur stratégie de contrôle. Cependant une étude comparative avec leur stratégie a montré que notre approche est plus robuste aux changements d'apparence et aux mouvements brusques de la cible.

Des investigations futures vont porter sur l'incorporation du zoom dans la stratégie ainsi qu'une modélisation plus complexe du mouvement de la cible. Nous nous intéresserons également à l'application de notre système dans un réseau de caméras à champs joints et disjoints.

Références

- [1] J.Ahmed, A.Ali and A.Khan, Stabilized active camera tracking system, *Journal of Real-Time Image Processing*, 2012.
- [2] J. Badri, Système de vision hybride : Modélisation et application au suivi haute résolution, *Thèse de l'Université Blaise Pascal - Clermont-Ferrand II*, 2008.
- [3] N.Bellotto, E.Sommerlade, B.Benfold and al., A distributed camera system for multi-resolution surveillance, *ACM/IEEE ICDCS*, 2009.
- [4] K.Bernardin, R.Stiefelhagen, Evaluating multiple object tracking performance : the CLEAR MOT metrics, *J. Image Video Process.*, 2008.
- [5] A.Biswas, P.Guha, A.Mukerjee and K.S.Venkatesh, Intrusion Detection and Tracking with Pan-Tilt Cameras, *IET ICVIE*, 2006.
- [6] H.C.Choi, U.Park, A.K.Jain and S.W.Lee, Face Tracking and Recognition at a Distance : A coaxial & concentric PTZ Camera System, *IEEE TCSVT*, 2011.
- [7] N.Dalal and B.Triggs Histograms of Oriented Gradients for Human Detection, *CVPR*, 2005.
- [8] I.Everts, N.Sebe and G.Jones, Cooperative Object Tracking with Multiple PTZ Cameras, *14th ICIAP*, 2007.
- [9] C.Guillot, M.Taron, P.Sayd, Q.C.Pham, C.Tilmant and J.M.Lavest, Background subtraction adapted to PTZ cameras by keypoint density estimation, *BMVC*, 2010.
- [10] S.Kang, J.Paik, A.Koschan, B.Abidi, and M.Abidi, Real-time video tracking using PTZ cameras, *IET International Conference on Quality Control by Artificial Vision*, 2003.
- [11] P.Kumar, A.Dick and T.S.Sheng, Real Time Target Tracking with Pan Tilt Zoom Camera, *DICTA*, 2009.
- [12] A.Mian, Realtime face detection and tracking using a single Pan, Tilt, Zoom camera, *23rd IVCNZ*, 2008.
- [13] R.Mottaghi and S.Payandeh, An overview of a probabilistic tracker for multiple cooperative tracking agents, *12th ICAR*, 2005.
- [14] P.Pérez, C.Hue, J.Vermaak and M.Gangnet Color-Based Probabilistic Tracking, *ECCV*, 2002.
- [15] F.Z.Qureshi and D.Terzopoulos, Planning ahead for PTZ camera assignment and handoff, *Third ACM/IEEE ICDCS*, 2009.
- [16] P.D.Z.Varcheie and G.A.Bilodeau, Adaptive Fuzzy Particle Filter Tracker for a PTZ Camera in an IP Surveillance System, *IEEE Transactions on Instrumentation and Measurement*, 2011.