Suivi visuel par filtrage particulaire. Application à l'interaction homme-robot

Ludovic Brèthes

N° d'ordre:

THÈSE

Présentée devant

l'Université Paul Sabatier de Toulouse

En vue de l'obtention du

Doctorat de l'Université Paul Sabatier de Toulouse
Spécialité Robotique

par

Ludovic Brèthes

Équipe d'accueil : LAAS-CNRS
École Doctorale : EDSYS

Titre de la thèse :

Suivi visuel par filtrage particulaire.
Application à l'interaction Homme-Robot

soutenue le 13 décembre 2005 devant la commission d'examen

Rapporteurs	M. Frédéric	Jurie	Chargé de Recherche CNRS, GRAVIR, Grenoble
	M. Patrick	Pérez	Directeur de Recherche IRISA/INRIA, Rennes
Examinateurs	M. Raja	Chatila	Directeur de Recherche CNRS, LAAS, Toulouse
	M. Patrice	Dalle	Professeur, UPS, IRIT (Toulouse)
Directeurs de thèse	M. Frédéric	Lerasle	MCf, UPS, LAAS, Toulouse
	M. Patrick	Danès	MCf, UPS, LAAS, Toulouse

Mieux vaut tard que jamais.
Libanios (314-394)

Remerciements

Avant tout, il convient de remercier Malik Ghallab, pour m'avoir accueilli dans son laboratoire.

Mes remerciements vont également à Raja Chatila, qui m'a accueilli au sein de son groupe de recherche et a présidé mon jury de soutenance.

Je remercie sincèrement MM. Frédéric Jurie et Patrick Pérez d'avoir accepté de rapporter cette thèse, ainsi que M. Patrice Dalle d'avoir eu la gentillesse de faire partie de mon jury de soutenance.

Ces trois années de thèse n'auraient pas pu se dérouler sans la présence et le soutien de Frédéric et Patrick mes co-directeurs de thèse, c'est pourquoi je les remercie chaleureusement de leur aide. Ils ont été très présents et ont su me conseiller et m'aiguiller lorsque j'en avais besoin. J'ai apprécié le travail dans cette équipe dynamique et soudée que nous avons formée.

Faire une thèse est une expérience qui demeure enrichissante à tout point de vue. Les trois ans passés au sein du groupe RIA m'auront beaucoup appris et cela aura aussi été un moyen de rencontrer de nombreuses personnes qui d'une façon ou d'une autre ont contribué à la réussite de ce travail. Je voudrais tout d'abord remercier Vincent avec qui j'ai partagé beaucoup de choses à commencer par le bureau, les cafés, les fous rires, le stress, ... un grand merci aussi à Thierry P et Aurélie qui ont été là depuis le début et m'ont aidé lorsque j'en avais besoin, sans oublier Sylvain A (merci pour le coup de pouce), Jérôme et Jean-Michel qui ont aussi été très présents ces dernières années. Naturellement, je remercie aussi tous les autres doctorants et stagiaires avec qui j'ai eu des contacts à la fois pour le travail et les loisirs, je pense particulièrement à Gabriel, Sylvain J, Mathias, Léonnard, Guillaume, Léo, Thomas, Olivier, Efrain, Aymeric, et bien d'autres encore ... Je voudrais aussi adresser mes remerciements à Paulo et Jean-Bernard qui m'ont beaucoup aidé à mon arrivée, ainsi qu'à Vivianne qui m'a gentillement accueilli dans son bureau et qui a toujours su se rendre disponible et a été de bon conseil, merci aussi à Thierry S.

Enfin je remercie de tout coeur toute ma famille pour son soutien inestimable et son encouragement permanent. Je remercie tout particulièrement Hélène pour m'avoir supporté même dans les moments les plus difficiles, c'est certainement grâce à elle que je suis arrivé jusqu'ici.

Table des matières

Chapitre 1

Introduction

1.1 Contexte général

Un défi de la Robotique aujourd'hui est sans doute celui du robot personnel. Cette perspective pose le problème essentiel de l'interaction et de la relation de l'homme au robot. Un objectif majeur est un robot mobile autonome qui navigue dans un environnement de grandes dimensions en présence de public. Lors de ses déplacements, le robot doit être capable de détecter et de prendre en compte de manière explicite la présence de personnes dans son voisinage pour les éviter ou leur céder le passage, le but étant de faciliter et de sécuriser leurs déplacements. Plus qu'un simple usager de l'environnement qu'il partage avec les humains, le robot doit en outre pouvoir interagir avec ces derniers par la reconnaissance de gestes élémentaires pour l'exécution de tâches décidées par eux : asservissement sur leurs déplacements, apprentissage supervisé, manipulation d'objets,...

Dans ce contexte, les travaux présentés ici portent plus spécifiquement sur la détection, le suivi de personnes et la reconnaissance de gestes élémentaires à partir du flot vidéo d'une caméra couleur embarquée sur un robot mobile. Le robot évolue dans des environnements d'intérieur (espaces ouverts connectés par un réseau de couloirs) *a priori* encombrés et sujets à des changements d'illumination. Il est alors opportun de gérer à chaque instant plusieurs hypothèses sur les paramètres à estimer et d'exploiter plusieurs sources de mesures.

Les fonctionnalités visuelles proposées doivent donc être tout à la fois robustes et simples. Les fonctions de suivi devront être robustes (i) aux mouvements *a priori* quelconques de la cible (translation, rotation, zoom), pouvant donner lieu à des occultations, (ii) aux conditions de prises de vue (changements d'illumination, scènes encombrées, caméra statique ou non, présence de plusieurs objets mobiles). La simplicité est relative à l'implémentation. Elle implique l'estimation d'un faible nombre de paramètres afin de répondre à des contraintes temporelles fortes.

Dans ce travail, nous nous limiterons à une analyse spatio-temporelle dans le plan image car elle est suffisamment pertinente du point de vue des modalités d'interaction

et pour permettre l'exploitation des résultats par d'autres modules de l'architecture du robot *e.g.* planification de trajectoire, surperviseur, planification de tâches, etc.

Enfin, ces approches 2D sont peu consommatrices de ressources CPU et ne compromettent donc pas l'exécution des autres fonctionnalités nécessaires à l'évolution autonome du robot. Par exemple, tout déplacement au long cours du robot impose une réactualisation de sa position dans la carte qu'il construit de l'environnement, et donc l'activation, entre autres, des modules d'acquisition et de traitement des données proprioceptives et extéroceptives.

1.2 Etat de l'art et positionnement de nos travaux

Le suivi visuel se définit conventionnellement comme le processus d'estimation des attributs de la cible dans le flot vidéo. Ces attributs sont relatifs à la position/dynamique, la forme (si elle est connue *a priori*) et l'apparence de la cible suivie. Son apparence est liée aux caractéristiques image que l'on peut extraire de la région d'intérêt associée.

Les approches traitant du suivi visuel de l'homme sont nombreuses dans la communauté et il serait présomptueux et illusoire de vouloir toutes les référencer. Dissocions d'emblée les approches 2D des approches 3D qui sortent du cadre de cette thèse pour les raisons évoquées précédemment. Pour plus de détails sur les approches 3D, le lecteur pourra se référer à [Moeslund et al., 2001, Sminchisescu, 2002].

Revenons un à un sur les attributs listés précédemment. Certains travaux considèrent naturellement l'attribut forme en exploitant la silhouette de tout ou partie des limbes corporels. Citons les travaux sur les contours actifs, notamment les *snakes* pour Isard et Blake [Isard et al., 1996b] ou les équations aux dérivées partielles pour Paragios et Deriche [Paragios et al., 2000]. Ces techniques permettent de prendre en compte les déformations d'une forme prototype (*template*). Des variantes consistent à définir un modèle statistique hiérarchique [Kervrann et al., 1996] voire à gérer plusieurs formes prototypes [Gavrila, 2000] pour estimer les déformations. Les méthodes d'analyse estimant à la fois le mouvement et la structure des membres nécessitent de résoudre un problème complexe et conduisent parfois à des solutions instables en présence de bruit. Des perturbations sont cependant incontournables lorsqu'aucune restriction ou hypothèse n'est faite quant au contexte de la prise d'images. La prise en compte de ces considérations et la volonté de simplifier au maximum la modélisation font que nos fonctions de suivi exploitant la forme privilégient une forme prototype grossière non déformable. Certains considérent des silhouettes plus ou moins approximatives [Isard et al., 1998b] ou des formes géométriques telles que des ellipses pour le suivi de visage [Birchfield, 1998, Schwerdt et al., 2000, Wagener et al., 2003, Rui et al., 2001] ou des ellipses et des cercles pour le suivi de la main [Bretzner et al., 2002].

Certains travaux exploitent l'apparence de la cible dans les différentes images du flot vidéo. Le *template* associé est représenté par des images propres [Blake et al., 1998a] ou régions d'intérêt [Jurie et al., 2002] en niveaux de gris, un sous-ensemble de points de contours [Huttenlocher et al., 1993], la phase de filtres en ondelettes [Jepson et al., 2001] ou plus largement des distributions de couleur, par exemple dans [Kawato et al., 2000,

Schwerdt et al., 2000, Wu et al., 2001]. L'information de couleur est ici modélisée par une ou plusieurs Gaussiennes [Wu et al., 2001, Thayananthan et al., 2003], une *Look-Up-Table* [Kawato et al., 2000] ou des histogrammes normalisés [Schwerdt et al., 2000, Comaniciu et al., 2000, Nummiaro et al., 2003, Pérez et al., 2004]. Ces histogrammes, connus pour leur robustesse aux occultations, leur confèrent un grand intérêt pour le suivi.

Ce modèle d'apparence peut résulter de connaissances *a priori* issues d'un apprentissage hors-ligne, par exemple de la couleur peau [Schwerdt et al., 2000, Pérez et al., 2004], d'images propres [Blake et al., 1998a] ou des variations de luminance associées aux mouvements image [Jurie et al., 2002]. Sans connaissance *a priori*, l'apparence de la cible doit être capturée dans le flot vidéo, par exemple par une segmentation sur le mouvement. Cette démarche est particulièrement adaptée aux applications de surveillance pour lesquelles l'arrière-plan est supposé plus ou moins stationnaire [Kawato et al., 2000, Haritaogly et al., 2000, Isard et al., 2001]. Elle permet de s'affranchir des fausses mesures liées à l'arrière-plan mais reste subordonnée à des hypothèses restrictives quant aux mouvements observés dans la scène. Plus globalement, l'initialisation du suivi (ou sa réinitialisation) dépend de la détection souvent *ad hoc* de la cible qui exploite le contexte général des prises de vue [Isard et al., 1998a, Nummiaro et al., 2003] afin d'en faciliter le processus.

En présence de scènes encombrées, une démarche plus générique est de mixer plusieurs informations dans le *template*, par exemple couleur et forme dans [Isard et al., 1998b, Birchfield, 1998, Davis et al., 2000, Wu et al., 2001, Bretzner et al., 2002], couleur et mouvement dans [Spengler et al., 2001, Pérez et al., 2004], couleur et son dans [Pérez et al., 2004], images d'intensité et mouvement dans [Blake et al., 1998a].

Les changements d'illumination, les occultations et mouvements de la cible impliquent la mise à jour de son apparence durant le processus de suivi. Cette mise à jour peut trivialement considérer l'apparence du *template* dans l'image précédente [Papanikolopoulos et al., 1993] mais une telle stratégie est sujette à dérives dans le temps voire échoue en présence d'occultations et de changements d'illumination brutaux. Un filtrage de l'apparence, sous l'hypothèse de variations lentes, est souvent appliqué, par exemple dans [Nguyen et al., 2001, Nummiaro et al., 2003].

Pour en simplifier la formulation, les approches dissocient souvent le filtrage de l'apparence du filtrage de la position et dynamique de la cible [Wu et al., 2001, Jepson et al., 2001, Nguyen et al., 2001, Vermaak et al., 2002a, Nummiaro et al., 2003]. Les approches reposent alors majoritairement sur le filtrage de Kalman [Blake et al., 1993, Huttenlocher et al., 1993, Papanikolopoulos et al., 1993, Schwerdt et al., 2000], le gradient du critère - *e.g. meanshift* [Comaniciu et al., 2003] ou ses variantes [Bradski, 1998, Chen et al., 2001]-, enfin les techniques de filtrage particulaire et ses variantes, citons par exemple [Isard et al., 1996a, Isard et al., 1998c, MacCormick et al., 2000, Nummiaro et al., 2003, Torma et al., 2003].

Les techniques de filtrage particulaire, abondamment référencées dans la littérature,

sont très adaptées à ce contexte. Il s'agit de méthodes de simulation séquentielles de type Monte Carlo permettant l'estimation du vecteur d'état d'un système Markovien non nécessairement linéaire soumis à des excitations aléatoires possiblement non Gaussiennes [Blake et al., 1998a]. Ce formalisme générique permet ainsi de s'affranchir de toute hypothèse restrictive quant aux distributions de probabilités entrant en jeu dans la caractérisation du problème.

Isard et Blake dans [Isard et al., 1996a] sont les premiers à exploiter le filtrage particulaire pour le suivi visuel. Ils introduisent l'algorithme bien connu de CONDENSATION - pour *Conditional Density Propagation* - qui s'appuie sur une fonction d'importance relative à la dynamique du système pour propager les particules dans l'espace d'état. Ceci confère à la CONDENSATION une structure prédiction/mise à jour comparable à celle du filtre de Kalman mais sans restriction à des modèles probabilistes Gaussiens. Les auteurs montrent les limites du filtre de Kalman sur des séquences de suivi de visages en milieux encombrés pour lesquels les ditributions de probabilité sont multi-modales. Dans ces exemples, l'estimation de la densité de probabilité *a posteriori* du vecteur d'état repose sur une fonction de mesure relative à la seule forme du visage et aux contours extraits dans l'image courante. Cet algorithme incontournable est souvent mis en œuvre et utilisé comme référence pour la comparaison avec d'autres stratégies de filtrage.

Ces mêmes auteurs introduisent dans [Isard et al., 1998c] une extension de la CONDENSATION (*mixed-state CONDENSATION*) afin de gérer des variables continues et discrètes dans le vecteur d'état. Dans leur application, une variable discrète indexe le modèle de dynamique associé à la cible parmi une bibliothèque donnée de modèles de mouvements canoniques apparents. L'évolution du paramètre discrêt entre les différents modèles est gérée par une matrice de transition contenant les probabilités de transition. Le suivi et la reconnaissance du modèle le plus vraisemblable s'effectuent simultanément. Les auteurs illustrent leur approche sur des séquences d'une main effectuant un dessin où les modèles de dynamique utilisés sont : courbe, tremblement et arrêt. Une application au suivi simultané de deux personnes est proposée dans [MacCormick et al., 1999]. La variable discrète permet ici de gérer les occultations mutuelles entre les deux sujets.

L'algorithme de ICONDENSATION proposé par Isard *et al.* dans [Isard et al., 1998a] est une autre extension de la CONDENSATION. Les auteurs définissent une stratégie de filtrage qui combine à la fois des informations visuelles bas-niveau et des mesures haut-niveau, où certaines particules sont échantillonnées selon une fonction d'importance relative à une mesure visuelle seule (détection de *blobs* couleurs) alors que d'autres suivent la dynamique du processus d'état sous-jacent. Au final, la densité de probabilité *a posteriori* du vecteur d'état est mise à jour par recalage d'un modèle de contour *i.e.* la silhouette d'une main. Cette stratégie de filtrage permet entre autre, la réinitialisation du filtre même si la détection de *blobs* peau est génératrice de fausses mesures pour des scènes encombrées.

Torma et Szepesvári dans [Torma et al., 2003] se placent le cas d'une fonction d'importance ne permettant d'échantillonner qu'une partie du vecteur d'état, typiquement

les parties du vecteur correspondant à l'innovation[1]. Ils soulignent le risque de « contradiction » pouvant survenir dès lors que le nouvel état est tiré indépendamment de son passé et proposent des extensions assurant de conserver dans cette situation la cohérence entre les parties historique et innovation des particules. Le principe est de sélectionner par des rééchantillonnages des paires d'innovation et d'historique qui ont une probabilité élevée de co-occurence. Dans l'article, trois algorithmes sont proposés. Le premier nommé HSSIR pour *History Sampling Sampling Importance Resampling* rééchantillonne les « parties innovations » selon leur probabilités d'occurence, puis, pour chacune des sous-particules ainsi obtenues, échantillonne un historique plausible. Une version « Rao-Blackwellisée » (RBHSSIR) est également développée afin de réduire la variance de l'estimateur en ne rééchantillonnant qu'un passé plausible pour chaque particule. Une dernière extension nommée RBSSHSSIR pour *Rao-Blackwellised History Sampling SIR* est une adaptation de RBHSSIR pour permettre la prise en compte de fonctions d'importance qui ne permettent d'échantillonner qu'une sous partie de l'innovation. Une expérimentation de suivi d'un objet artificiel combinant une détection de couleur et la mesure de forme montre un gain en précision et taux de réussite des stratégies proposées par les auteurs comparé à la CONDENSATION et au filtre SIR basique.

MacCormick et Isard dans [MacCormick et al., 2000] exploitent l'échantillonnage partitionné introduit dans [MacCormick et al., 1999] pour l'étendre au suivi d'objets articulés tels que la main. Le principe est de diviser l'espace d'état en plusieurs « partitions », puis d'appliquer séquentiellement la dynamique sur chaque partition préalablement à un rééchantillonnage pondéré adapté. Le but est de réduire le nombre de particules nécéssaires, notamment dans un contexte de grande dimension.

Rui et Chen dans [Rui et al., 2001] s'appuient sur le développement du filtrage particulaire *unscented* dans la communauté de traitement du signal ([Merwe et al., 2000]) pour l'appliquer dans un contexte de localisation de locuteur et de suivi visuel. Cette stratégie permet d'incorporer l'observation courante dans la propagation des particules au moyen d'une fonction d'importance Gaussienne associée à chaque particule et mise à jour par un filtre de Kalman *unscented*. La prise en compte à la fois de la dynamique du processus d'état et de l'observation à l'instant courant positionne au mieux les particules dans l'espace d'état. La cible à suivre correspond ici au contour d'un visage, et est modélisée par une ellipse. Les vraisemblances des particules sont basées sur la mesure des distances entre les ellipses associées et les points de contours de l'image. Les auteurs comparent leur approche à la CONDENSATION et montrent l'avantage de cette stratégie sur des exemples de suivi réels.

Li et Zhang dans [Li et al., 2002] réalisent le suivi d'une main modélisée par son contour afin de comparer la CONDENSATION au *filtre particulaire unscented* et au *filtre particulaire Kalman*. Ces deux dernières stratégies apparaissent plus précises que la CONDENSATION en terme d'erreur quadratique moyenne, cependant, les temps de calcul prohibitifs associés au filtre particulaire *unscented* conduisne les auteurs à

[1] Le sens donné par les auteurs au vocable « innovation » doit être distingué du terme français relatif à l'erreur de prédiction sur la mesure, dont la traduction anglo-saxonne serait *residual*.

conclure que le *filtre particulaire Kalman* est la meilleure stratégie dans un contexte de suivi visuel de mains.

Un autre avantage connu du filtrage particulaire, en dehors de sa grande généricité, est de pouvoir combiner/fusionner aisément différentes sources de mesures. Malgré ce constat, la fusion de données par filtrage particulaire nous semble assez peu exploitée et souvent confinée à un nombre restreint de primitives visuelles. Néanmoins, certains travaux proposent des fonctions d'importance permettant l'obtention de performances intéressantes.

Pérez *et al.* dans [Pérez et al., 2002] introduisent une nouvelle technique de suivi par Monte Carlo basée sur l'apparence de la cible. La cible à suivre est ici caractérisée par un *template* d'histogrammes indépendants normalisés de la couleur relative aux pixels inclus dans une région rectangulaire englobant celle-ci. La fonction de mesure est définie par une mesure de similarité basée sur la distance de Bhattacharyya entre histogrammes couleur. Cette nouvelle approche probabiliste est comparée à l'algorithme *meanshift* [Comaniciu et al., 2003] qui utilise le même attribut visuel et montre des résultats similaires. Les auteurs proposent ensuite trois contributions. La première définit une nouvelle fonction de mesure qui combine les distributions de couleur contenues dans deux régions de la cible afin d'ajouter une contrainte spatiale entre les distributions de couleur. Le gain de cette stratégie de mesure par rapport à la précédente est montrée sur une séquence réelle de suivi. La deuxième contribution concerne la prise en compte du modèle de couleur du fond dans le calcul des vraisemblances. Dans un contexte où la caméra est immobile et où une image du fond est disponible, les vraisemblances sont alors obtenues à partir de la différence entre la distance colorimétrique cible/référence et la distance colorimétrique cible/fond. Enfin, la dernière contribution concerne le suivi multi-objets. Les auteurs tirent ici parti de la capacité des filtres particulaires à capturer les multi-modalités. Le vecteur d'état est alors constitué de la concaténation des vecteurs d'états des différentes cibles et une méthode d'exclusion est mise en œuvre dans le calcul des vraisemblances dès lors qu'il y a un recouvrement (occultation) entre les cibles. La méthode est finalement expérimentée sur une séquence où deux personnes se croisent ; le *tracker* réussit à les suivre en conservant le bon ordre de profondeur et sans les confondre.

Nummiaro *et al.* dans [Nummiaro et al., 2003] exploitent aussi la CONDENSATION ainsi qu'une mesure basée sur la distribution de couleur, mais la cible à suivre est ici modélisée par une ellipse. Durant le suivi, ces distributions de couleur sont mises à jour afin de gérer les changements d'apparence de la cible, typiquement la rotation d'une personne sur elle-même. La fonction de mesure est définie par une mesure de similarité basée sur la distance de Bhattacharyya entre histogrammes couleur. L'approche proposée est alors comparée avec un algorithme de *meanshift* [Comaniciu et al., 2003] à partir de séquences-test. Ce dernier est plus rapide et plus précis mais pour des distributions possiblement multi-modales, la CONDENSATION est plus robuste aux fausses mesures. Les auteurs montrent que le nombre de particules influence directement la précision du système.

Isard et MacCormick dans [Isard et al., 2001] définissent un *tracker multi-blobs* basé sur un filtrage par CONDENSATION. Leur travail se situe plus particulièrement au niveau de la définition de la fonction de mesure exploitée dans le filtre. Le fond de la scène est modélisé par un mélange de Gaussiennes capturant à la fois les informations sur la couleur et sur le gradient contenus aux différents points d'une grille de taille fixe sous-échantillonnant l'image. Un autre mélange de Gaussiennes du même type capture les informations associées au premier plan de l'image typiquement des personnes. Les poids des particules sont alors obtenus en calculant le rapport entre les deux mélanges de gaussiennes dans une région candidate, ici la projection dans le plan image d'un cylindre 3D modélisant la personne. L'algorithme nommé BraMBLe pour *Bayesian Multiple-Blob Tracker*, permet de suivre un nombre de personnes *a priori* inconnu et variable. Les expérimentations présentées dans l'article montrent les performances de cette approche, qui suit sans difficulté trois personnes. Un problème apparait cependant lorsque deux personnes se croisent, du fait de la commutation erronée des labels qui leur sont associées.

MacCormick et Blake dans [MacCormick et al., 1999] proposent une solution pour suivre plusieurs objets dont le modèle ne permet pas de les différentier, *e.g.* contours de la tête. Ils introduisent dans cet article deux contributions majeures. Tout d'abord, une méthode d'exclusion mutuelle empêche une même mesure de contribuer aux calculs des vraisemblances de différentes cibles. Ensuite, l'échantillonnage partitionné pour les méthodes de Monte Carlo est implémenté, ce qui permet, par une exploration plus efficace de l'espace d'état du système, de diminuer significativement le nombre de particules nécessaires au suivi.

Outre les travaux tels que ceux déja mentionnés de Isard *et al.* dans [Isard et al., 1998a], d'autres approches permettent de condidérer des attributs visuels variés pour le suivi.

Spengler et Schiele dans [Spengler et al., 2001] mixent deux cartes de salience afin de segmenter respectivement les régions de couleur peau ou mobiles (mouvement inter-images). Chaque fonction de mesure est définie par un mélange de Gaussiennes caractérisé à l'aide d'un algorithme EM (*Expectation Maximization*). La densité de probabilité *a posteriori* de l'état est estimée par CONDENSATION, où la fonction de mesure globale est une somme pondérée des fonctions de mesure précédentes. Les auteurs comparent alors avec une estimation par maximum de vraisemblance et concluent que la CONDENSATION est plus performante lorsque les densités estimées sont multi-modales.

Bretzner *et al.* dans [Bretzner et al., 2002] définissent un modèle hiérarchique grossier d'une main à partir d'ellipses et de cercle, respectivement pour les doigts et la paume. La fonction de mesure est définie par une différence quadratique entre les mélanges de Gaussiennes relatives au modèle et à l'image. La finalité est ici de suivre et reconnaître dans le flot vidéo, par CONDENSATION, la configuration la plus vraisemblable de la main. Les gestes analysés sont grossièrement fronto-parallèles à la caméra. Chaque pixel est affecté d'une probabilité d'appartenance à la classe peau. Les auteurs montrent alors que la fusion, dans la fonction de mesure, des vraisemblances relatives

au modèle et de couleur réduit sensiblement les décrochages du filtre en présence de scènes encombrées.

Pérez *et al.* ont largement abordé le problème de fusion de données dans [Pérez et al., 2004]. Ils proposent un algorithme de filtrage hiérarchisé (*hierarchical sampling*) dans lequel deux mesures sont consommées : (1) une mesure intermittente pour éventuellement positionner efficacement les particules, (2) une mesure persistente pour pondérer ces dernières et donc mettre à jour la densité *a posteriori*. Les mesures intermittentes tirent partie de modules de détection soit visuel (mouvement inter-images), soit acoustique. Enfin, les mesures persistentes sont relatives au seul attribut couleur. Dans notre contexte, la connaissance *a priori* des cibles suivies permet de considérer en plus la forme dans nos fonctions de mesure.

Tous ces travaux, bien que liés par le même cadre applicatif, décrivent des fonctions de suivi très variées en termes de stratégies de filtrage et de mesures. Hélas, ils proposent assez peu d'éléments de comparaison et illustrent, souvent sur quelques séquences-clés, le comportement qualitatif de leurs filtres.

Lichtenauer *et al.* dans [Lichtenauer et al., 2004] traitent de l'influence de la fonction de vraisemblance sur les performances d'un suivi par filtre particulaire. Dans cet article, les auteurs observent le comportement du suivi par CONDENSATION dans une image synthétique en fonction de la forme de la fonction de vraisemblance. Trois scénarii sont envisagés : (1) suivi d'un objet seul, (2) suivi d'objets multiples et (3) réinitialisation (objet perdu). Ils montrent que les paramètres définissant la vraisemblance, typiquement la covariance dans le lien état-mesure, doivent être adaptés au contexte de suivi considéré.

Peu de travaux proposent d'évaluer plus en détail les différentes stratégies de filtrage et de mesure. Il nous semble pertinent de les comparer et de les évaluer plus largement dans un contexte robotique.

Plus généralement, il nous semble intéressant de proposer des filtres qui combinent ou fusionnent plus largement les mesures visuelles dans leurs fonctions d'importance et de mesure. A terme, nous prévoyons, par exemple, d'intégrer dans nos filtres des informations non visuelles. De plus, nous souhaitons également évaluer quelques-unes des nombreuses stratégies de filtrage proposées dans la littérature de façon à déterminer les associations de primitives visuelles et d'algorithmes d'estimation qui répondent au mieux aux modalités d'interaction envisagées pour notre robot. Le filtrage particulaire offre ici un cadre suffisamment générique pour l'implémentation des différentes fonctionnalités de suivi visuel associées. Nous tirons partie des plateformes mobiles du groupe Robotique et Intelligence Artificielle (RIA) pour définir des modalités réalistes d'interaction entre celles-ci et l'humain dans un environnement *a priori* quelconque.

Ces modalités décrivent une interaction entre l'homme et son guide, ici le robot, dans un musée, car ces travaux s'inscrivent dans le cadre d'une collaboration du groupe RIA avec la Cité de l'Espace de Toulouse. L'objectif est de voir sur site le robot nommé Rackham utilisé pour ce projet naviguer et interagir avec les visiteurs de la Cité sur la base des fonctionnalités intégrées. Dans un contexte similaire, ces travaux s'intègrent

aussi dans le projet européen COGNIRON (Cognitive Robot Companion) coordonné par le LAAS-CNRS. L'objectif central de ce projet est de conférer des capacités cognitives à des robots à travers l'étude et le développement de méthodes et de technologies pour la perception, l'interprétation, le raisonnement et l'apprentissage en interaction avec l'homme. Nos contributions pour ce projet se situent donc sur les aspects interaction visuelle Homme-Robot.

Enfin, l'ensemble des fonctionnalités visuelles que nous proposons doivent obéir à des contraintes temps-réelles réalistes avec le contexte applicatif décrit précédemment.

1.3 Organisation du manuscrit

Après ce chapitre introductif, les quatre chapitres suivants s'organisent comme suit :

Le **chapitre** 2 rappelle quelques généralités sur le filtrage particulaire et détaille l'algorithme générique. Les difficultés ainsi que les leviers sur lesquels il est possible d'agir lors de la définition d'un filtre sont ensuite abordés. Différentes stratégies simples d'échantillonnage sont alors présentées. Elles reposent sur des fonctions d'importance relatives à la dynamique ou aux mesures. Enfin, la stratégie récursive optimale, et les algorithmes permettant de s'en approcher sont discutés.

Le **chapitre** 3 spécifie, pour notre contexte applicatif et le formalisme précédent, des mesures visuelles reposant sur des attributs de couleur, forme ou mouvement inter-images de la cible observée. Ces mesures visuelles sont prises en compte dans le modèle de mesure et/ou la fonction d'importance des filtres. En présence d'environnements encombrés, il est alors judicieux de combiner ou fusionner plusieurs mesures de natures différentes. Des stratégies combinant ou fusionnant tout ou partie des attributs précités sont donc envisagées. Les diverses fonctions de mesure ou d'importance associées sont finalement évaluées et comparées sur une base d'images acquises depuis le robot et représentatives des scènes rencontrées lors de sa navigation.

Le **chapitre** 4 décrit des modalités complémentaires d'interaction pour notre robot-guide. Différentes fonctionnalités de suivi, adaptées à chacun des scénarii, sont alors proposées et mises en œuvre. Elles se définissent en termes de mesures visuelles et de stratégies de filtrage dont les choix sont argumentés. Chaque fonctionnalité visuelle est ensuite confrontée à des contextes suffisamment variés et réalistes pour une modalité donnée d'interaction afin d'en évaluer les performances en termes de robustesse, précision et temps d'exécution. Ces évaluations permettent de valider les fonctionnalités de suivi pour chacune des modalités d'interaction.

Le **chapitre** 5 présente enfin des modalités de suivi et reconnaissance de gestes symboliques d'une main humaine. Le recalage du *template* et la sélection de la configuration de la main la plus vraisemblable s'effectuent dans un seul et unique processus de filtrage particulaire, ce qui permet d'obtenir des temps de calcul compatibles avec les applications visées. Comme précédemment, deux stratégies spécifiques de filtrage particulaire sont comparées. La seconde, proche du cas optimal, est originale car jamais

appliquée pour le suivi visuel. Des résultats de suivi et de reconnaissance sont alors décrits et discutés pour ces modalités. Celles-ci doivent permettre aux usagers partageant l'environnement du robot de le commander gestuellement à partir d'un vocabulaire qui sort du cadre de cette thèse.

Le bilan de ces travaux ainsi que les perspectives envisagées sont présentés en conclusion.

Chapitre 2

Estimation par filtrage particulaire

L'estimation de l'état d'un système dynamique à partir d'observations bruitées et éventuellement incomplètes est un problème central dans de nombreuses applications. On peut citer, par exemple, l'analyse de signaux radars, la localisation 3D d'objets en robotique, la modélisation de données financières, le traitement de la parole ou, notamment dans notre contexte, le suivi visuel d'entités.

Généralement, le système est modélisé par une chaîne de Markov cachée à temps discret, dont l'état vit dans un espace continu ou bien admet des valeurs discrètes. L'objectif est alors d'estimer à chaque instant la loi de ce processus à partir de réalisations du processus d'observation. Un tel système est entièrement décrit par la distribution du processus d'état à l'instant initial, un modèle d'évolution de l'état ainsi qu'un modèle de mesure reliant l'état à l'observation. De manière générale, toute l'information qui peut être inférée sur l'état disposant des mesures jusqu'à un instant quelconque k est capturée dans sa loi conjointe *a posteriori*, *i.e.* dans sa distribution conjointe depuis l'instant initial jusqu'à l'instant k conditionnellement à la connaissance des mesures jusqu'à k. Très souvent, seule la loi marginale est recherchée, au sens où il s'agit uniquement de caractériser la distribution *a posteriori* de l'état à l'instant k.

Dans ce travail, nous nous intéressons au problème du filtrage, qui consiste à établir la distribution *a posteriori* – conjointe ou marginale – de manière récursive.

Sous certaines hypothèses – *e.g.* espace d'état discret et fini, système décrit par une représentation d'état soumise à des excitations aléatoires Gaussiennes –, il est possible de propager temporellement la loi *a posteriori* exacte ou une approximation de ses premiers moments. Cependant, dans un contexte de suivi visuel tel que le nôtre, il peut être nécessaire de recourir à une alternative "sous-optimale" plus générique, permettant d'approximer la loi *a posteriori* y compris si elle est multimodale, si le lien état-mesure est non linéaire, si les bruits sont non Gaussiens, etc. Le filtrage particulaire constitue une telle alternative, du fait qu'il permet l'estimation approchée de l'état – discret ou continu – de tout système dynamique Markovien, indépendamment de toute hypothèse sur la nature des bruits.

Les filtres particulaires sont des méthodes séquentielles de Monte Carlo consistant à approcher la densité *a posteriori* au moyen d'un ensemble de mesures ponctuelles

– ou « particules » – pondérées. Les méthodes de Monte Carlo sont apparues dans les années 50 avec l'article de Metropolis et Ulam [Metropolis et al., 1949] suivi d'autres travaux comme [Metropolis et al., 1953] ou [Hammersley et al., 1954]. Puis, dans les années 60-70, les faibles puissances de calcul ainsi que la dégénérescence des algorithmes due à leur implémentation basée sur un échantillonnage pondéré séquentiel brut ont conduit à une perte d'intérêt pour ces méthodes inutilisables en pratique. Seuls quelques développements ponctuels tels que [Handschin et al., 1970] ou [Akashi et al., 1977] ont continué à explorer ces idées. Plus récemment, dans les années 90, les méthodes de Monte Carlo ont été de nouveau explorées, avec l'introduction du rééchantillonnage par Gordon *et al.* [Gordon et al., 1993] qui permet de limiter la dégénérescence des algorithmes basés sur l'échantillonnage pondéré séquentiel. Cette contribution majeure associée à l'augmentation importante des capacités de calcul a eu un impact déterminant dans la communauté Traitement du Signal en rendant les filtres particulaires utilisables en pratique pour la première fois.

Depuis, les activités de recherche dans ce domaine ont énormément augmenté [Doucet et al., 2001a] conduisant à de nombreuses contributions améliorant l'efficacité des filtres particulaires. Ces méthodes ont été largement exploitées dans divers domaines comme par exemple le traitement du signal [Gustafsson et al., 2002], [Hue et al., 2000], le traitement de la parole [Vermaak et al., 2002a], la robotique mobile [Kwok et al., 2004], la modélisation de données financières [Pitt et al., 2001], ou le suivi visuel [Isard et al., 1998a], [Pérez et al., 2004].

Ce chapitre introduit tout d'abord les méthodes séquentielles de Monte Carlo, puis aborde leurs difficultés ainsi que les « leviers » sur lesquels il est possible d'agir lors de la définition d'un filtre. Un algorithme générique de filtrage particulaire s'en suit. Ensuite, plusieurs stratégies d'échantillonnage sont évoquées, ainsi que des mécanismes permettant d'améliorer le filtrage. La dernière partie discute la stratégie récursive optimale, et les algorithmes permettant de s'en approcher.

2.1 Généralités

Dans cette section, nous présentons les outils et méthodes mis en œuvre dans l'estimation de l'état d'un système dynamique par filtrage particulaire. Après avoir présenté le formalisme et le principe des méthodes de Monte Carlo, nous abordons les difficultés liées à ce type d'estimateur ainsi que les techniques permettant de pallier ces problèmes. Ensuite, nous décrivons l'algorithme générique de filtrage particulaire qui peut en être déduit et pour finir nous détaillons le cas optimal de filtrage.

Afin de simplifier les notations, nous noterons une variable aléatoire de la même façon que sa réalisation. Dans le cas continu, toute probabilité $\mathbb{P}(X \in \mathrm{d}x)$ sera supposée de la forme $p(x)\mathrm{d}x$. De même, dans le cas discret, la distribution $\mathbb{P}(X = x)$ sera notée $p(x)$. Enfin, nous commettrons souvent l'abus de langage consistant à confondre une distribution de probabilité et sa densité.

2.1.1 Formalisation du problème

A Modélisation

On considère un système stochastique dynamique qui est le siège d'un *processus d'état*, caché, lequel est indirectement observé *via* un *processus d'observation*. Les représentations à temps discret de ces deux processus aléatoires sont respectivement désignées par $\{x_k\}_{k\in\mathbb{N}}$ et $\{z_k\}_{k\in\mathbb{N}^*}$, avec $x_k \in \mathbb{R}^{n_x}$, $z_k \in \mathbb{R}^{n_z}$.

Soient $x_{0:k} \triangleq \{x_0, \ldots, x_k\}$ et $z_{1:k} \triangleq \{z_1, \ldots, z_k\}$. Le processus $\{x_k\}$ admet une distribution initiale caractérisée par $p(x_0)$, et est supposé Markovien de loi de transition $p(x_k|x_{0:k-1}) = p(x_k|x_{k-1})$. Les observations sont supposéees indépendantes conditionnellement au processus d'état, et leur distribution ne dépend que de l'état au même instant, ce qui se traduit par

$$\begin{gathered} p(z_{1:k}|x_{0:k}) = p(z_k|x_{0:k})p(z_{1:k-1}|x_{0:k}), \\ \text{où } p(z_k|x_{0:k}) = p(z_k|x_{0:k}, z_{1:k-1}) \text{ satisfait } p(z_k|x_{0:k}) = p(z_k|x_k). \end{gathered} \tag{2.1}$$

La densité de probabilité $p(z_k|x_k)$ de z_k conditionnellement à x_k permet naturellement de quantifier la *vraisemblance* de x_k par rapport à z_k. En outre, le vecteur de mesure à un instant donné et le vecteur d'état à l'instant suivant obéissent à la propriété d'indépendance conditionnelle

$$\begin{gathered} p(x_k, z_{1:k-1}|x_{0:k-1}) = p(x_k|x_{0:k-1})p(z_{1:k-1}|x_{0:k-1}), \\ \text{où } p(x_k|x_{0:k-1}) = p(x_k|x_{0:k-1}, z_{1:k-1}) \text{ satisfait } p(x_k|x_{0:k-1}) = p(x_k|x_{k-1}). \end{gathered} \tag{2.2}$$

À titre d'exemple, la modélisation (2.2)–(2.1) recouvre la représentation d'état

$$\begin{cases} x_{k+1} = f(x_k) + w_k \\ z_k = h(x_k) + v_k \end{cases} \tag{2.3}$$

d'un système non linéaire Markovien soumis à des bruits de dynamique w_k et de mesure v_k blancs, mutuellement indépendants, et indépendants de la condition initiale.

B Le filtrage Bayésien optimal et sa solution récursive exacte

Basé sur le caractère Markovien du système et la modélisation précédente, le filtrage Bayésien optimal consiste en le calcul récursif de la densité conjointe *a posteriori* $p(x_{0:k}|z_{1:k})$ ou de sa densité marginale – également désignée ci-après « distribution de filtrage » – $p(x_k|z_{1:k})$.

L'application de la règle de Bayes permet l'écriture récursive de $p(x_{0:k}|z_{1:k})$ en fonction de $p(x_{0:k-1}|z_{1:k-1})$ sous la forme

$$p(x_{0:k}|z_{1:k}) = \frac{p(z_k|x_{0:k}, z_{1:k-1})p(x_{0:k}|z_{1:k-1})}{p(z_k|z_{1:k-1})} \tag{2.4}$$

$$\text{avec} \quad p(x_{0:k}|z_{1:k-1}) = p(x_k|x_{0:k-1}, z_{1:k-1})p(x_{0:k-1}|z_{1:k-1}). \tag{2.5}$$

Les propriétés d'indépendance (2.2) et (2.1) permettent de simplifier (2.4)–(2.5) en

$$\begin{aligned} p(x_{0:k}|z_{1:k}) &= \frac{p(z_k|x_k)p(x_k|x_{k-1})}{p(z_k|z_{1:k-1})} p(x_{0:k-1}|z_{1:k-1}) \\ &\propto p(z_k|x_k)p(x_k|x_{k-1})p(x_{0:k-1}|z_{1:k-1}). \end{aligned} \tag{2.6}$$

La marginalisation de (2.5) selon $x_{0:k-1}$ conduit à l'équation de Chapman-Kolgomorov, qui exprime la densité de prédiction $p(x_k|z_{1:k-1})$ à l'instant k en fonction de la densité de filtrage $p(x_{k-1}|z_{1:k-1})$ à l'instant précédent $k-1$. En invoquant la propriété (2.2), cette équation s'écrit

$$p(x_k|z_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}. \tag{2.7}$$

Similairement à (2.4), la distribution de filtrage à l'instant k est obtenue à partir de la distribution de prédiction en ce même instant par une mise à jour prenant en compte l'observation z_k, soit

$$\begin{aligned} p(x_k|z_{1:k}) &= \frac{p(z_k|x_k)}{p(z_k|z_{1:k-1})} p(x_k|z_{1:k-1}) \\ &\propto p(z_k|x_k)p(x_k|z_{1:k-1}). \end{aligned} \tag{2.8}$$

La constante de normalisation $p(z_k|z_{1:k-1})$ intervenant dans (2.6) et (2.8) s'exprime en fonction de la densité de prédiction et du lien état-mesure sous la forme

$$p(z_k|z_{1:k-1}) = \int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k. \tag{2.9}$$

Dès lors que la distribution *a posteriori* $p(x_k|z_{1:k})$ est connue, un estimé du vecteur d'état peut être défini à partir de la donnée d'un critère d'optimalité. Ainsi, l'estimé du maximum *a posteriori* et l'estimé du minimum d'erreur quadratique moyenne[1] s'écrivent respectivement $[\hat{x}_{k|k}]_{\text{MAP}} = \arg\max_{x_k} p(x_k|z_{1:k})$ et $[\hat{x}_{k|k}]_{\text{MMSE}} = \mathbb{E}_{p(.|z_{1:k})}(x_k)$. Plus généralement, toute quantité de la forme $\mathbb{E}_{p(.|z_{1:k})}(\phi(x_k))$ peut être évaluée, avec $\phi(.)$ une fonction intégrable de $\mathbb{R}^{n_x}$ à valeurs dans $\mathbb{R}$, permettant ainsi le calcul de la covariance *a posteriori*, etc.

La détermination récursive de la distribution *a posteriori* ne peut être effectuée analytiquement que dans quelques situations très particulières. Ainsi, dans le cas d'un système linéaire soumis à des bruits de dynamique et de mesure Gaussiens, on montre qu'il s'agit d'une Gaussienne, dont les deux premiers moments peuvent être propagés au moyen du filtre de Kalman. Cependant, pour des modèles plus complexes, la solution exacte du problème ne peut pas être déterminée, *e.g.* du fait de l'impossibilité de calculer l'expression analytique de (2.9) dans (2.4)–(2.5) ou (2.7)–(2.8). Il est alors nécessaire d'utiliser des méthodes approchées parfois appelées algorithmes Bayésiens sous-optimaux [Arulampalam et al., 2002].

[1] *Minimum Mean-Square Error* (MMSE) estimate.

Ainsi, le filtre de Kalman étendu (FKE) – ou "*Extended Kalman Filter*" – permet la détermination approchée de la moyenne et de la covariance *a posteriori* du vecteur d'état par la linéarisation au premier ordre des équations de dynamique et de mesure, respectivement autour du dernier estimé et de la dernière prédiction disponibles. Cependant, le fait de négliger tous les termes d'orde supérieur ou égal à 2 dans les développements de Taylor des non-linéarités conduit souvent à une performance médiocre de ce filtre, voire à sa divergence.

Le filtre de Kalman "Unscented" (FKU) – ou "*Unscented Kalman Filter*" – récemment proposé par Julier et Uhlmann [Julier et al., 1996] permet une précision plus accrue. Reposant sur un échantillonnage déterministe, au moyen de "σ-points" judicieusement sélectionnés, de la distribution *a posteriori* à l'instant précédent ainsi que des bruits de dynamique et de mesure, il conduit, *via* l'utilisation de la "transformée Unscented", à une approximation des deux premiers moments de la distribution *a posteriori* à l'instant courant. Celle-ci est valable jusqu'au deuxième ordre des développements en série de Taylor des non-linéarités, sans augmentation de complexité comparativement au FKE. L'approximation tient jusqu'au troisième ordre si les distributions des bruits et du processus d'état sont assimilées à des Gaussiennes.

Afin de s'affranchir de l'hypothèse de Gaussianité de la loi *a posteriori*, des approximations par mélanges de Gaussiennes ont également été développées [Alspach et al., 1972]. Dans notre contexte non-linéaire, multi-modal et non-Gaussien, nous préfèrerons des algorithmes de filtrage particulaire, ou – « méthodes séquentielles de Monte Carlo » – qui approximent des lois de probabilité continues au moyen de distributions ponctuelles. Ces dernières méthodes présentent quelques similitudes avec les méthodes de filtrage par maillage de l'espace d'état –"*grid-based methods*" –, qui approximent la densité *a posteriori* par une somme de mesures de Dirac admettant pour support une grille figée[2]. Cependant, elles se différencient par le fait que les distributions ponctuelles élémentaires constituant l'approximation particulaire sont centrées sur une grille qui évolue de manière stochastique, afin de permettre une exploration adaptative des zones « pertinentes » de l'espace d'état.

2.1.2 Filtrage particulaire

A Méthodes d'approximation de Monte Carlo

A-1 Principe Les méthodes de Monte Carlo permettent d'approximer une distribution de probabilité continue $p(x)$ quelconque au moyen d'une distribution discrète de la forme

$$\hat{p}_N(x) \triangleq \sum_{i=1}^{N} w^{(i)} \delta(x - x^{(i)}), \text{ avec } \sum_{i=1}^{N} w^{(i)} = 1, \tag{2.10}$$

[2]Lorsque l'espace d'état est fini et borné, de telles méthodes sont exactes dès lors que la grille recouvre l'ensemble des valeurs possibles.

de sorte que simuler un nombre selon la loi $p(x)$ revient à sélectionner un échantillon – ou « particule » – $x^{(i)}$ avec la probabilité – ou « poids » – $w^{(i)}$. Dès lors, toute intégrale

$$p(\Phi) \triangleq \int \Phi(x)p(x)\mathrm{d}x \tag{2.11}$$

relative à l'espérance de l'image par une fonction $\Phi(.)$ d'une variable aléatoire se distribuant selon $p(x)$ peut être approchée par

$$\hat{p}_N(\Phi) \triangleq \int \Phi(x)\hat{p}_N(x)\mathrm{d}x = \sum_{i=1}^{N} w^{(i)}\Phi(x^{(i)}). \tag{2.12}$$

A-2 Échantillonnage « idéal » Dans le cas où on sait échantillonner $p(x)$, une première approche consiste à définir $x^{(1)}, \ldots, x^{(N)}$ comme des variables aléatoires indépendantes identiquement distribuées (i.i.d.) selon $p(x)$, ou, de manière équivalente, comme des nombres indépendamment simulés selon $p(x)$, et à les affecter de poids identiques. Sous ces conditions, que l'on note

$$x^{(i)} \sim p(x),\ w^{(i)} = \frac{1}{N}, \tag{2.13}$$

on montre que l'estimateur

$$p_N(\Phi) = \frac{1}{N}\sum_{i=1}^{N} \Phi(x^{(i)}) \tag{2.14}$$

est non biaisé, et, d'après la loi forte des grands nombres, converge presque sûrement – *i.e.* avec une probabilité de 1– vers $p(\Phi)$ lorsque N tend vers l'infini. Si la variance $\sigma_\Phi^2 = \mathrm{Var}(\Phi(x))$ est finie, alors la variance de $p_N(\Phi)$ est égale à $\frac{\sigma_\Phi^2}{N}$. La convergence en loi de l'erreur d'estimation est alors garantie pour toute fonction $\Phi(.)$ continue bornée par le théorème central limite

$$\frac{\sqrt{N}}{\sigma_\Phi}(p_N(\Phi) - p(\Phi)) \xrightarrow{loi} \mathcal{N}(0,1), \tag{2.15}$$

d'où peuvent être extraits des intervalles de confiance de $p(\Phi)$ centrés sur $\hat{p}_N(\Phi)$ lorsque N tend vers l'infini [Campillo, 2005, Millet, 2005]. On note que la vitesse de convergence $\frac{1}{\sqrt{N}}$ ne dépend pas de la dimension de l'espace dans lequel vit x.

A-3 Échantillonnage préférentiel Lorsqu'il n'est pas possible ou souhaité de tirer des échantillons de $p(x)$, la stratégie précédente est inapplicable. Une alternative est l'échantillonnage préférentiel – ou "importance sampling" / « échantillonnage pondéré » –, qui consiste à sélectionner $x^{(1)}, \ldots, x^{(N)}$ selon une distribution erronée, puis à compenser numériquement cette opération dans les poids $w^{(1)}, \ldots, w^{(N)}$.

Plus précisément, supposons que les particules $x^{(1)}, \ldots, x^{(N)}$ soient échantillonnées de manière indépendante selon une densité –ou « loi d'importance »– $q(x)$ telle que $p(x) > 0$ implique $q(x) > 0$. L'intégrale $p(\Phi)$ définie en (2.11) s'écrit également

$$p(\Phi) = \int \Phi(x) \frac{p(x)}{q(x)} q(x) \mathrm{d}x, \tag{2.16}$$

de sorte qu'elle peut être approximée par la somme $\frac{1}{N}\Sigma_{i=1}^{N} w^{*(i)} \Phi(x^{(i)})$, avec $w^{*(i)} = \frac{p(x^{(i)})}{q(x^{(i)})}$. Cependant, l'estimateur $\hat{p}_N(\Phi)$ defini en (2.12) avec $w^{(i)} = \frac{1}{N} w^{*(i)}$ ne satisfait pas la condition de normalisation $\sum_{i=1}^{N} w^{(i)} = 1$ figurant en (2.11). C'est pourquoi il faut poser $w^{(i)} = \frac{w^{*(i)}}{\sum_{j=1}^{N} w^{*(j)}}$. En résumé, l'ensemble de particules pondérées $\{x^{(i)}, w^{(i)}\}$ constitue une description cohérente de $p(x)$ dès lors que

$$x^{(i)} \sim q(x), \; w^{(i)} \propto \frac{p(x^{(i)})}{q(x^{(i)})}, \tag{2.17}$$

préalablement à une étape de normalisation telle que $\sum_{i=1}^{N} w^{(i)} = 1$.

Pour les mêmes raisons que précédemment, l'estimateur $\hat{p}_N(\Phi)$ ainsi construit est asymptotiquement non biaisé et converge presque sûrement vers $p(\Phi)$ quelle que soit la distribution d'importance $q(x)$ dont le support recouvre celui de $p(x)$. Sa variance est finie seulement si l'espérance $\mathbb{E}_{q(.)}(\Phi^2(x)\frac{p^2(x)}{q^2(x)}) = \int \Phi^2(x) \frac{p^2(x)}{q(x)} \mathrm{d}x$ est finie. On retrouve un théorème central limite semblable à (2.15), qui permet ici aussi d'encadrer $p(\Phi)$ dans un intervalle centré sur $\hat{p}_N(\Phi)$ lorsque $N \longrightarrow +\infty$.

L'intérêt de l'échantillonnage préférentiel est double [Millet, 2005, Moulines, 2002]. D'une part, il peut permettre une approximation de $p(\Phi)$ plus efficace, par la réduction de la variance de son estimateur. Ceci implique que le rapport $\frac{p(x)}{q(x)}$ soit borné, sous peine que les poids $w^{(i)}$ varient beaucoup et ne soient élevés que pour un nombre restreint de particules. D'autre part, comme cela a été indiqué plus haut, il permet de ne pas échantillonner selon $p(x)$. Ceci est particulièrement intéressant lorsque la loi $p(x)$ n'est connue qu'à une constante de normalisation près.

Il convient toutefois de mentionner que ces propriétés nécessitent un effort calculatoire plus important au niveau de l'évaluation des poids.

À titre d'exemple, supposons que l'on souhaite simuler une loi normale $\mathcal{N}_T(\mu, \Sigma)$ de moyenne μ et de covariance Σ tronquée sur le segment $[\mu - T; \mu + T]$. Une possibilité consiste à sélectionner de manière équiprobable une particule parmi N i.i.d. selon $\mathcal{N}_T(\mu, \Sigma)$. En outre, définir N particules $x^{(i)}$ i.i.d. selon la loi uniforme de support $[\mu - T; \mu + T]$, et leur affecter des poids $w^{(i)}$ proportionnels à $\exp(-\frac{1}{2}(x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu))$ permet également l'obtention d'une description particulaire cohérente.

B Application à la problématique du filtrage

B-1 Échantillonnage préférentiel Dans ce contexte, il s'agit donc d'établir une représentation particulaire de la loi conjointe *a posteriori* $p(x_{0:k}|z_{1:k})$ ou de la loi margi-

nale $p(x_k|z_{1:k})$. Du fait que ces densités ne sont généralement connues qu'à une constante de normalisation près – cf. (2.6) ou (2.8) – l'échantillonnage préférentiel est invoqué.

On choisit pour cela une distribution d'importance $q(x_{0:k}|z_{1:k})$ pouvant être facilement échantillonnée et dont le support inclut le support de $p(x_{0:k}|z_{1:k})$, *i.e.*

$$\forall x_{0:k} \in (\mathbb{R}^{n_x})^{k+1},\ p(x_{0:k}|z_{1:k}) > 0 \Rightarrow q(x_{0:k}|z_{1:k}) > 0. \tag{2.18}$$

Conformément à (2.10), on peut écrire

$$p(x_{0:k}|z_{1:k}) \approx \hat{p}_N(x_{0:k}|z_{1:k}) = \sum_{i=1}^{N} w_{0:k}^{(i)} \delta(x_{0:k} - x_{0:k}^{(i)}), \tag{2.19}$$

dès lors que les échantillons $x_{0:k}^{(i)}$ sont i.i.d. et affectés de poids $w_{0:k}^{(i)}$ selon [3]

$$x_{0:k}^{(i)} \sim q(x_{0:k}|z_{1:k}),\ w_{0:k}^{(i)} \propto \frac{p(x_{0:k}^{(i)}|z_{1:k})}{q(x_{0:k}^{(i)}|z_{1:k})}, \sum_{i=1}^{N} w_{0:k}^{(i)} = 1. \tag{2.20}$$

L'intégrale

$$I(f) = \int f(x_{0:k}) p(x_{0:k}|z_{1:k}) dx_{0:k} \tag{2.21}$$

relative à l'espérance *a posteriori* d'une fonction $f(.)$ de la trajectoire d'état depuis l'instant 0 jusqu'à l'instant k est par conséquent approximée au moyen de l'estimateur

$$\hat{I}_N(f) = \sum_{i=1}^{N} w_{0:k}^{(i)} f(x_{0:k}^{(i)}) \tag{2.22}$$

dont les propriétés ont déjà été évoquées au §2.1.2–A.

Notons qu'une marginalisation triviale de (2.19) par rapport à $x_{0:k-1}$ permet de déduire l'approximation particulaire de la densité de filtrage, qui s'écrit

$$p(x_k|z_{1:k}) \approx \hat{p}_N(x_k|z_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)}),\ \text{où}\ w_k^{(i)} = w_{0:k}^{(i)}. \tag{2.23}$$

B-2 Échantillonnage pondéré séquentiel et méthode séquentielle de Monte Carlo L'échantillonnage pondéré combiné à la méthode d'intégration de Monte Carlo permet d'estimer la densité *a posteriori*. Cette estimation doit cependant être formulée de manière récursive, *i.e.* l'approximation particulaire de $p(x_{0:k}|z_{1:k})$ doit être déterminée à partir du nuage de particules pondérées approchant $p(x_{0:k-1}|z_{1:k-1})$ ainsi que de la nouvelle observation z_k. Si on choisit une fonction d'importance « causale », satisfaisant

$$\forall k' \geq k,\ q(x_{0:k}|z_{1:k'}) = q(x_{0:k}|z_{1:k}), \tag{2.24}$$

[3] Notations : $x_{0:k}^{(i)} = i^{me}$ particule, réalisation de la trajectoire $x_{0:k}$
$w_{0:k}^{(i)}$ = poids, scalaire, associé à cette particule.

alors, du fait que $q(x_{0:k}|z_{1:k}) = q(x_k|x_{0:k-1}, z_{1:k})q(x_{0:k-1}|z_{1:k})$ s'écrit

$$q(x_{0:k}|z_{1:k}) = q(x_k|x_{0:k-1}, z_{1:k})q(x_{0:k-1}|z_{1:k-1}), \tag{2.25}$$

chaque nouvelle $i^{\text{ème}}$ particule $x_{0:k}^{(i)} \sim q(x_{0:k}|z_{1:k})$ peut être définie comme l'« augmentation » de la $i^{\text{ème}}$ particule $x_{0:k-1}^{(i)} \sim q(x_{0:k-1}|z_{1:k-1})$ à l'instant précédent, par un nouvel état $x_k^{(i)}$ sélectionné selon $q(x_k|x_{0:k-1}^{(i)}, z_{1:k})$.

La substitution de (2.25) et (2.6) dans (2.20) conduit à l'équation récursive de mise à jour des poids d'importance :

$$\begin{aligned} w_k^{(i)} &\propto \frac{p(x_{0:k}^{(i)}|z_{1:k})}{q(x_{0:k}^{(i)}|z_{1:k})} \\ &\propto \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})p(x_{0:k-1}^{(i)}|z_{1:k-1})}{q(x_k^{(i)}|x_{0:k-1}^{(i)}, z_{1:k})q(x_{0:k-1}^{(i)}|z_{1:k-1})} \\ &\propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{0:k-1}^{(i)}, z_{1:k})}. \end{aligned} \tag{2.26}$$

Si, de plus, la fonction d'importance satisfait $q(x_k|x_{0:k-1}, z_{1:k}) = q(x_k|x_{k-1}, z_k)$, les poids (2.26) dépendent uniquement de l'état précédent et de l'observation courante, au sens où

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_k)}. \tag{2.27}$$

Dans notre contexte de suivi, nous supposons que la fonction d'importance est de la forme $q(x_k|x_{k-1}, z_k)$ et nous ne nous intéressons qu'à l'estimation de la loi de filtrage $p(x_k|z_{1:k})$.

Ces développements permettent de définir l'algorithme d'échantillonnage pondéré séquentiel ("*Sequential Importance Sampling*", SIS) résumé Table 2.1, qui construit récursivement un nuage de particules pondérées approchant la loi de filtrage $p(x_k|z_{1:k})$ à l'instant k. Chaque particule $x_{k-1}^{(i)}$ est « propagée » selon la fonction d'importance, puis les poids sont mis à jour selon (2.27) préalablement à leur normalisation. Rappelons que l'approximation (2.19) de la loi conjointe *a posteriori* peut immédiatement être déduite en définissant $x_{0:k}^{(i)} = \{x_0^{(i)}, \ldots, x_k^{(i)}\}$ et en utilisant la propriété (2.23).

Cet algorithme très simple d'estimation de la loi de filtrage $p(x_k|z_{1:k})$ possède l'avantage d'être parallélisable. Toutefois, sa nature récursive soulève certains problèmes. Dans la section suivante, nous abordons les difficultés liées à cette stratégie de filtrage, puis discutons des points sensibles et des méthodes qui peuvent contribuer à une amélioration de l'efficacité du filtre.

$\left[\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^N\right] = \text{SIS}\left[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1}^N, z_k\right]$

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
3: **FIN SI**
4: **SI** $k \geq 1$ **ALORS**
5: **POUR** $i = 1, \ldots, N$, **FAIRE**
6: « Propager » la particule $x_{k-1}^{(i)}$ en simulant de manière indépendante $x_k^{(i)} \sim q(x_k | x_{k-1}^{(i)}, z_k)$
7: Mettre à jour le poids $w_k^{(i)}$ selon l'équation

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_k)}$$

8: **FIN POUR**
9: Normaliser les poids d'importance

$$w_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^N w_k^{(j)}}$$

de sorte que $\sum_{i=1}^N w_k^{(i)} = 1$
10: Le nuage $\{x_k^{(i)}, w_k^{(i)}\}_{i=1\ldots N}$ permet d'approcher la loi de filtrage par

$$p(x_k|z_{1:k}) \simeq \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$$

11: **FIN SI**

TAB. 2.1 – Algorithme de filtrage par échantillonnage pondéré séquentiel (SIS)

2.1.3 Difficultés et leviers

On a vu comment construire un estimateur récursif de la distribution conjointe *a posteriori* $p(x_{0:k}|z_{1:k})$ ainsi que de la loi de filtrage $p(x_k|z_{1:k})$. La précision de l'estimateur peut être qualifiée au moyen de sa variance. Dans cette section, nous nous proposons de discuter certains problèmes relatifs à ce critère, et présentons des méthodes et techniques permettant leur minimisation.

A Dégénérescence de l'algorithme

En raison de la nature récursive de l'algorithme SIS, la variance inconditionnelle des poids dans le temps augmente [Kong et al., 1994], induisant une dégradation de la précision de l'estimation. Ceci se traduit par le phénomène de *dégénérescence* du nuage, au sens où après un certain nombre d'étapes de récursion, la plupart des particules sont affectées d'un poids normalisé négligeable. Il se peut alors que des ressources importantes soient nécessaires aux calculs liés à leur évolution, pour une contribution finalement insignifiante dans l'approximation de $p(x_k|z_{1:k})$. La dégénérescence de l'algorithme est d'autant plus sensible que la distribution d'importance $q(x_{0:k}|z_{1:k})$ diffère de $p(x_{0:k}|z_{1:k})$, qui correspondrait au cas équipondéré.

Une mesure du phénomène de dégénérescence est la *taille efficace du N-échantillon [Kong et al., 1994]*, ci-après désignée par N_{eff}. Ce critère fait intervenir le rapport entre la variance de l'estimateur reposant sur un échantillonnage préférentiel selon $q(x_{0:k}|z_{1:k})$ et celle de l'estimateur obtenu en échantillonnant selon la loi *a posteriori* $p(x_{0:k}|z_{1:k})$. Il est défini par

$$\begin{aligned} N_{eff} &= \frac{N}{\frac{\mathrm{Var}_{q(.|z_{1:k})}\hat{I}_N(\Phi(x_{0:k}))}{\mathrm{Var}_{p(.|z_{1:k})}I_N(\Phi(x_{0:k}))}} \\ &= \frac{N}{1+\mathrm{Var}_{q(.|z_{1:k})}\frac{p(x_{0:k}^{(i)}|z_{1:k})}{q(x_{0:k}^{(i)}|z_{1:k})}}. \end{aligned} \tag{2.28}$$

L'expression de N_{eff} ne peut pas être calculée directement, mais une estimation est donnée par :

$$\widehat{N_{eff}} = \frac{1}{\sum_{i=1}^{N}(w_k^{(i)})^2} \tag{2.29}$$

où $w_k^{(i)}$, $i = 1, \ldots, N$ désignent les poids normalisés obtenus par l'équation (2.26). Les valeurs maximale $N_{eff} = N$ et minimale $N_{eff} = 1$ de N_{eff} correspondent respectivement à l'équipondération des particules – *i.e.* $\forall i = 1, \ldots, N$, $w_k^{(i)} = \frac{1}{N}$ – et au cas de dégénérescence extrême : $\exists j \in \{1, \ldots, N\}$ tel que $w_k^{(j)} = 1$ et $w_k^{(i)} = 0$ pour tout $i \neq j$.

Une autre méthode d'approximation du N_{eff} proposée par [Carpenter et al., 1999] met en œuvre un calcul basé sur une méthode de Monte Carlo. D'autres critères de mesure de la dégénérescence des poids peuvent être envisagés, *e.g.* basés sur des mesures d'entropie, etc. Toutefois, on ne retient généralement que la première méthode, qui est plus simple et moins coûteuse en temps de calcul.

Afin de pallier partiellement ce problème de dégénérescence et d'augmenter l'efficacité d'un filtre particulaire, plusieurs stratégies peuvent être mises en œuvre.

B Choix de la fonction d'importance – Stratégie récursive optimale

La fonction d'importance définit la stratégie d'exploration de l'espace d'état par les particules. Or, le positionnement de chaque particule conditionne son poids *via* sa vraisemblance par rapport à l'observation et la compatibilité de son historique vis à vis de la dynamique du système. On comprend donc que la fonction d'importance ait une influence particulière sur la dispersion des poids et par conséquent sur l'efficacité du filtre.

Un échantillonnage selon la distribution *a posteriori* $p(x_{0:k-1}|z_{1:k})$ conduirait à un nuage de particules équipondéré et par conséquent, comme indiqué précédemment, à $N_{eff} = N$. Il n'est évidemment pas possible de poser $q(x_{0:k-1}|z_{1:k}) = p(x_{0:k-1}|z_{1:k})$, car la condition (2.24) permettant d'exprimer la fonction d'importance sous la forme récursive (2.25) ne tiendrait plus. Bien que (2.25) implique l'augmentation de la variance inconditionnelle des poids dans le temps, il demeure néanmoins nécessaire de limiter la dégénerescence de l'algorithme de filtrage.

Ainsi, la fonction d'importance est dite *optimale* si elle permet de minimiser la variance des poids *conditionnellement à* $z_{1:k}$ *et* $x_{0:k-1}^{(i)}$. D'après [Doucet et al., 2000], celle-ci satisfait

$$\begin{aligned} q^*(x_k|x_{k-1}^{(i)}, z_k) &= p(x_k|x_{k-1}^{(i)}, z_k) \\ &= \frac{p(z_k|x_k, x_{k-1}^{(i)})p(x_k|x_{k-1}^{(i)})}{p(z_k|x_{k-1}^{(i)})}. \end{aligned} \tag{2.30}$$

En substituant cette équation dans (2.26), les pondérations des particules deviennent

$$\begin{aligned} w_k^{*(i)} &\propto w_{k-1}^{*(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{p(x_k|x_{k-1}^{(i)}, z_k)} \\ &\propto w_{k-1}^{*(i)} p(z_k|x_{k-1}^{(i)}). \end{aligned} \tag{2.31}$$

Une propriété remarquable de (2.31) est que le poids $w_k^{*(i)}$ associé à chaque particule $x_k^{(i)}$ ne dépend que de la particule prédécesseur $x_{k-1}^{(i)}$. Il s'en suit que la variance des poids conditionnellement à $z_{1:k}$ et $x_{0:k-1}^{(i)}$ est nulle.

Il advient toutefois qu'on ne puisse pas échantillonner selon $p(x_k|x_{k-1}^{(i)}, z_k)$ ou que la vraisemblance

$$p(z_k|x_{k-1}^{(i)}) = \int p(z_k|x_k)p(x_k|x_{k-1}^{(i)})\mathrm{d}x_k \tag{2.32}$$

ne puisse pas être exprimée analytiquement. c'est notamment le cas dans le contexte de suivi visuel. Néanmoins, même si seule une approximation de $p(x_k|x_{k-1}^{(i)}, z_k)$ et/ou de $p(z_k|x_{k-1}^{(i)})$ est disponible, il demeure possible de se rapprocher du cas optimal, cf. §2.3.

C Introduction du rééchantillonnage – Vers un algorithme générique de filtrage particulaire

L'utilisation de l'échantillonnage pondéré séquentiel ne suffit pas à estimer convenablement la loi *a posteriori*, du fait que la dégénérescence des poids ne permet pas une couverture pertinente de l'espace d'état. Néanmoins, comme mentionné au §2.1.2–A, une distribution peut être approximée par différents nuages de particules pondérées. Sur ce principe, une étape de rééchantillonnage peut permettre de limiter le phénomène de dégénérescence par une duplication des particules fortement pondérées au détriment de celles, affectées d'un poids faible, qui disparaissent. Pour cela, l'ensemble de particules pondérées $\{x_k^{(j)}, w_k^{(j)}\}$ est transformé en un ensemble de mesures aléatoires $\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)} = 1/N\}$ de poids uniformes en sélectionnant, avec remise, les N nouvelles particules $\tilde{x}_k^{(i)}$ dans l'ensemble $\{x_k^{(j)}\}$ selon la règle $\mathbb{P}(\tilde{x}_k^{(i)} = x_k^{(j)}) = w_k^{(j)}$, de sorte que l'étape de rééchantillonnage peut être résumée en

$$\begin{aligned} \tilde{x}_k^{(i)} &= x_k^{(j)} \text{ avec une probabilité } w_k^{(j)} \\ \tilde{w}_k^{(i)} &= \frac{1}{N}. \end{aligned} \tag{2.33}$$

Pour être cohérent, *i.e.* pour qu'il permette l'obtention de particules $\tilde{x}_k^{(i)}$ i.i.d. selon $\sum_{j=1}^{N} w_k^{(j)} \delta(x_k - x_k^{(j)})$, un rééchantillonnage doit dupliquer chaque particule $x_k^{(j)}$ un nombre N_j de fois proportionnel à son poids $w_k^{(j)}$. Les méthodes proposées dans la littérature assurent qu'en moyenne on a bien $\mathbb{E}(N_j) = N w_k^{(j)}$, toutefois elles ajoutent systématiquement une variance – dite « de Monte Carlo » – sur N_j qui a pour conséquence d'introduire une imprécision supplémentaire sur la distribution estimée. C'est pourquoi nous avons donc choisi d'utiliser la méthode à variance de Monte Carlo minimale dite de « rééchantillonnage systématique » [Kitagawa, 1996], qui est en outre simple à implémenter et de complexité en $O(N)$. Cet algorithme résumé Table 2.2 assure que le nombre de duplications de chaque particule $x_k^{(j)}$ ne diffère pas de $N w_k^{(j)}$ de plus de 1.

$\left[\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)}\}_{i=1}^N\right] = RESAMPLE\left[\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^N\right]$

Initialiser la somme cumulée des poids (SCP) : $c_1 = w_k^{(1)}$
POUR $i = 2, \ldots, N$, **FAIRE**
  Construire SCP : $c_i = c_{i-1} + w_k^{(i)}$
FIN POUR
Démarrer au début de SCP : $i = 1$
Tirer un point de départ : $u_1 \sim \mathcal{U}[0, N^{-1}]$
POUR $j = 1, \ldots, N$, **FAIRE**
  Se déplacer le long de SCP : $u_j = u_1 + (j-1)N^{-1}$
  TANT QUE $u_j > c_i$, **FAIRE**
    $i = i + 1$
  FIN TANT QUE
  Recopier la particule : $\tilde{x}_k^{(j)} = x_k^{(i)}$
  Affecter le poids : $w_k^{(j)} = N^{-1}$
FIN POUR

TAB. 2.2 – Algorithme dit de « rééchantillonnage systématique » [Kitagawa, 1996]

Intégré à l'algorithme SIS, le rééchantillonnage constitue donc un moyen de limiter la dégénérescence des poids d'importance. Il faut néanmoins limiter autant que possible le nombre de rééchantillonnages, d'une part en raison de la variance de Monte Carlo introduite mais aussi afin de limiter le phénomène d'*appauvrissement* du nuage résultant d'une répétition éventuellement trop importante des particules de poids $w_k^{(j)}$ élevés. Cette perte de diversité dans l'exploration de l'espace d'état est d'autant plus importante que la dynamique du système est peu bruitée. Une solution consiste à ne rééchantillonner que lorsque l'estimée $\widehat{N_{eff}}$ établi en (2.29) de la taille efficace du N-échantillon se situe en dessous d'un certain seuil $\widehat{N_{eff}} < N_s$.

Un algorithme générique de filtrage particulaire, nommé SIR pour *Sampling Importance Resampling*, découle de ces considérations, cf. Table 2.3. Tous les filtres particulaires peuvent être considérés comme une instance de SIR et ne diffèrent que par le choix de la fonction d'importance et par la stratégie de rééchantillonnage mise en œuvre. Il convient toutefois de signaler, pour une meilleure estimation de l'intégrale $I(f)$ définie

en (2.21) au moyen de $\hat{I}_N(f)$ dans (2.22), qu'il est préférable de ne pas procéder au rééchantillonnage à chaque instant et que l'évaluation de (2.22) doit être effectuée de préférence avant le rééchantillonnage.

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N = \text{SIR}([\{x_{k-1}^{(i)}, w_{k-1}^{(i)},\}]_{i=1}^N, z_k)$

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
3: **FIN SI**
4: **SI** $k \geq 1$ **ALORS**
5: **POUR** $i = 1, \ldots, N$, **FAIRE**
6: « Propager » la particule $x_{k-1}^{(i)}$ en simulant de manière indépendante

$$x_k^{(i)} \sim q(x_k | x_{k-1}^{(i)}, z_k) \tag{2.34}$$

7: Mettre à jour le poids $w_k^{(i)}$ selon l'équation

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_k)} \tag{2.35}$$

préalablement à une étape de normalisation assurant que $\sum_{i=1}^N w_k^{(i)} = 1$
8: **FIN POUR**
9: Le nuage $\{x_k^{(i)}, w_k^{(i)}\}_{i=1\ldots N}$ permet d'approcher la loi de filtrage par

$$p(x_k|z_{1:k}) \simeq \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$$

10: De manière systématique où dès lors que $\frac{1}{\sum_{i=1}^N (w_k^{(i)})^2} < \textit{seuil}$, rééchantillonner $\{x_k^{(i)}, w_k^{(i)}\}$ selon $P(\tilde{x}_k^{(i)} = x_k^{(j)}) = w_k^{(j)}$, de façon à obtenir un ensemble de particules pondérées $\{\tilde{x}_k^{(i)}, \frac{1}{N}\}$ tel que $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ et $\frac{1}{N} \sum_{i=1}^N \delta(x_k - \tilde{x}_k^{(i)})$ approximent $p(x_k|z_{1:k})$. Affecter $x_k^{(i)}$ et $w_k^{(i)}$ avec $\tilde{x}_k^{(i)}$ et $\frac{1}{N}$
11: **FIN SI**

TAB. 2.3 – Algorithme générique de filtrage particulaire (SIR)

Signalons enfin que l'introduction d'une étape de rééchantillonnage dans l'algorithme SIR, outre le fait qu'elle remet en cause sa parallélisation de manière aussi immédiate que pour le SIS, complexifie considérablement les preuves relatives à sa convergence. En effet, les particules étant statistiquement dépendantes à l'issue du rééchantillonnage, les résultats classiques de convergence des méthodes de Monte Carlo, qui reposent sur l'hypothèse "i.i.d." –cf. §2.1.2 –A– ne permettent pas de conclure. Le lecteur intéressé est invité à consulter [Crisan et al., 2002]. La thèse de MacCormick [MacCormick, 2000] contient également une reformalisation accessible d'une preuve de Del Moral relative à ce problème.

D Autres stratégies de rééchantillonnage et d'échantillonnage

D-1 Rééchantillonnage pondéré Le rééchantillonnage pondéré [MacCormick, 2000] permet, disposant d'une approximation particulaire $\sum_{j=1}^{N} w_k^{(j)} \delta(x_k - x_k^{(j)})$ d'une distribution donnée de la variable x_k, de construire un autre nuage pondéré $\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)}\}$ représentant cette distribution, le positionnement des nouvelles particules $\tilde{x}_k^{(i)}$ étant effectué dans les zones de l'espace d'état où une fonction $g(.)$ donnée admet des valeurs élevées. Alors que le rééchantillonnage systématique de [Kitagawa, 1996] utilise les poids $w_k^{(j)}$ pour redistribuer les particules $x_k^{(j)}$, cette nouvelle approche définit l'ensemble $\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)}\}$ par

$$\begin{aligned} \tilde{x}_k^{(i)} &= x_k^{(j)} \text{ avec une probabilité } \rho_k^{(j)} = \frac{g(x_k^{(j)})}{\sum_{l=1}^{N} g(x_k^{(l)})}; \\ \tilde{w}_k^{(i)} &\propto \frac{w_k^{(j)}}{\rho_k^{(j)}}. \end{aligned} \tag{2.36}$$

La fonction $g(.)$ est supposée continue et à valeurs strictement positives. Le nuage pondéré obtenu est effectivement une représentation cohérente de la distribution considérée dès lors qu'il existe une fonction $f(.)$, continue et à valeurs dans $\mathbb{R}_+^*$, telle que $w_k^{(j)}$ s'écrit

$$w_k^{(j)} = \frac{f(x_k^{(j)})}{\sum_{r=1}^{N} f(x_k^{(r)})}. \tag{2.37}$$

D-2 Échantillonnage partitionné Supposons que l'espace d'état soit défini comme le produit cartésien $X = X_1 \times \ldots \times X_M$ et que la dynamique du système s'écrive, avec $x_k = \left((x_k^1)', \ldots, (x_k^M)'\right)', x_k^m \in X_m, m = 1, \ldots, M$,

$$p(x_k|x_{k-1}) = \int \tilde{d}_1(\xi_1|x_{k-1}) \tilde{d}_2(\xi_2|\xi_1) \ldots \tilde{d}_M(x_k|\xi_{M-1}) d\xi_1 \ldots d\xi_{M-1}, \tag{2.38}$$

où, les vecteurs ξ_m étant partitionnés[4] comme x_k en $\xi_m = \left((\xi_m^1)', \ldots, (\xi_m^M)'\right)'$, $m = 1, \ldots, M$, les fonctions $\tilde{d}_1(.|.), \ldots, \tilde{d}_M(.|.)$ satisfont

$$\begin{aligned} \tilde{d}_1(\xi_1|x_{k-1}) &= d_1(\xi_1^1, \ldots, \xi_1^M|x_{k-1}), \\ \tilde{d}_2(\xi_2|\xi_1) &= d_2(\xi_2^2, \ldots, \xi_2^M|\xi_1).\delta(\xi_2^1 - \xi_1^1), \\ \tilde{d}_3(\xi_3|\xi_2) &= d_3(\xi_3^3, \ldots, \xi_3^M|\xi_2).\delta(\xi_3^1 - \xi_2^1).\delta(\xi_3^2 - \xi_2^2), \\ &\vdots \\ \tilde{d}_M(x_k|\xi_{M-1}) &= d_M(x_k^M|\xi_{M-1}).\delta(x_k^1 - \xi_{M-1}^1).\delta(x_k^2 - \xi_{M-1}^2). \ldots .\delta(x_k^{M-1} - \xi_{M-1}^{M-1}). \end{aligned} \tag{2.39}$$

[4] Les indices des vecteurs ξ_m n'ont cependant aucun lien avec le temps, contrairement à l'indice k de $x_k \ldots$

En d'autres termes, la dynamique *a priori* du système se présente comme l'application successive –*i.e.* la convolution– de dynamiques élémentaires $\tilde{d}_1(.|.), \ldots, \tilde{d}_m(.|.), \ldots, \tilde{d}_M(.|.)$, chacune se focalisant sur un nombre plus restreint de constituants du vecteur d'état au fur et à mesure que $m \longrightarrow M$.

Supposons également que l'on dispose de fonctions $g_1(.), \ldots, g_m(.), \ldots, g_{M-1}(.)$ définies sur l'espace d'état $\mathbb{R}^{n_x}$ telles que chaque $g_m(.)$ admette un mode dans les zones de forte probabilité *a posteriori* de la $m^{\text{ième}}$ composante du vecteur d'état. Ces fonctions sont généralement définies à partir de la mesure.

Sous ces hypothèses, on montre que l'efficacité du filtre peut être significativement améliorée si l'approximation particulaire de la distribution *a posteriori* du vecteur d'état est déterminée au moyen des étapes suivantes [MacCormick, 2000, §7.6] :

1. Disposant du nuage pondéré $\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}$ représentant $p(x_{k-1}|z_{1:k-1})$, échantillonner de manière indépendante $(x_k^{1(i)}, \xi_1^{2(i)}, \ldots, \xi_1^{M(i)}) \sim d_1(\xi_1^1, \ldots, \xi_1^M | x_{k-1}^{(i)})$, $i = 1, \ldots, N$, de sorte que $\{(x_k^{1(i)}, \xi_1^{2(i)}, \ldots, \xi_1^{M(i)}), w_{k-1}^{(i)}\}$ approxime $\int \tilde{d}_1(\xi_1|x_{k-1}) p(x_{k-1}|z_{1:k-1}) dx_{k-1}$;

2. Obtenir l'approximation particulaire $\{(\tilde{x}_k^{1(i)}, \tilde{\xi}_1^{2(i)}, \ldots, \tilde{\xi}_1^{M(i)}), \tau_1^{(i)}\}$ "équivalente" à $\{(x_k^{1(i)}, \xi_1^{2(i)}, \ldots, \xi_1^{M(i)}), w_{k-1}^{(i)}\}$, en effectuant sur ce dernier nuage un rééchantillonnage pondéré guidé par l'ensemble des poids $\{g_1(x_k^{1(i)}, \xi_1^{2(i)}, \ldots, \xi_1^{M(i)})\}$; renommer $\{(\tilde{x}_k^{1(i)}, \tilde{\xi}_1^{2(i)}, \ldots, \tilde{\xi}_1^{M(i)}), \tau_1^{(i)}\}$ en $\{(x_k^{1(i)}, \xi_1^{2(i)}, \ldots, \xi_1^{M(i)}), \tau_1^{(i)}\}$;

3. Échantillonner, de manière indépendante, $(x_k^{2(i)}, \xi_2^{3(i)}, \ldots, \xi_2^{M(i)}) \sim d_2(\xi_2^2, \xi_2^3, \ldots, \xi_2^M | x_k^{1(i)}, \xi_1^{2(i)}, \ldots, \xi_1^{M(i)})$, $i = 1, \ldots, N$, de sorte que $\{(x_k^{1(i)}, x_k^{2(i)}, \xi_2^{3(i)}, \ldots, \xi_2^{M(i)}), \tau_1^{(i)}\}$ approxime $\int \tilde{d}_2(\xi_2|\xi_1) [\int \tilde{d}_1(\xi_1|x_{k-1}) p(x_{k-1}|z_{1:k-1}) dx_{k-1}] d\xi_1$;

4. Obtenir l'approximation particulaire $\{(\tilde{x}_k^{1(i)}, \tilde{x}_k^{2(i)}, \tilde{\xi}_2^{3(i)}, \ldots, \tilde{\xi}_2^{M(i)}), \tau_2^{(i)}\}$ "équivalente" à $\{(x_k^{1(i)}, x_k^{2(i)}, \xi_2^{3(i)}, \ldots, \xi_2^{M(i)}), \tau_1^{(i)}\}$, en effectuant sur ce dernier nuage un rééchantillonnage pondéré guidé par l'ensemble des poids $\{g_2(x_k^{1(i)}, x_k^{2(i)}, \xi_2^{3(i)}, \ldots, \xi_2^{M(i)})\}$; renommer $\{(\tilde{x}_k^{1(i)}, \tilde{x}_k^{2(i)}, \tilde{\xi}_2^{3(i)}, \ldots, \tilde{\xi}_2^{M(i)}), \tau_2^{(i)}\}$ en $\{(x_k^{1(i)}, x_k^{2(i)}, \xi_2^{3(i)}, \ldots, \xi_2^{M(i)}), \tau_2^{(i)}\}$;

   $\vdots$

$(2M-1)$. Échantillonner $x_k^{M(i)} \sim d_M(x_k^M | x_k^{1(i)}, x_k^{2(i)}, \ldots, x_k^{M-1(i)}, \xi_{M-1}^{M(i)})$ de façon à obtenir un ensemble de particules pondérées, noté $\{x_k^{(i)} = (x_k^{1(i)}, \ldots, x_k^{M-1(i)}, x_k^{M(i)}), \tilde{w}_{k-1}^{(i)} = \tau_{M-1}^{(i)}\}$, représentant $p(x_k|z_{1:k-1})$; confronter cet ensemble à la mesure z_k de façon que la mise à jour des poids $w_k^{(i)} \propto \tilde{w}_{k-1}^{(i)} p(z_k|x_k^{(i)})$ conduise à l'approximation particulaire $p(x_k|z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$.

Dans de nombreux cas, chacune des fonctions $d_i(.|.)$ intervenant dans les dynamiques

élémentaires $\tilde{d}_i(.|.)$ satisfait en outre

$$d_1(\xi_1^1,\ldots,\xi_1^M|x_{k-1}) = p(\xi_1^1|x_{k-1}^1)\prod_{r=2}^{M}\delta(\xi_1^r - x_{k-1}^r),$$

$$\forall m \in \{2,\ldots,M-1\},\ d_m(\xi_m^m,\ldots,\xi_m^M|\xi_{m-1}) = p(\xi_m^m|\xi_{m-1}^m)\prod_{r=m+1}^{M}\delta(\xi_m^r - \xi_{m-1}^r), \quad (2.40)$$

$$d_M(x_k^M|\xi_{M-1}) = p(x_k^M|\xi_{M-1}^M), \quad (2.41)$$

de sorte que $p(x_k|x_{k-1})$ devient

$$p(x_k|x_{k-1}) = \prod_{m=1}^{M} p(x_k^m|x_{k-1}^m). \quad (2.42)$$

De même, la vraisemblance $p(z_k|x_k)$ de $x_k = \left((x_k^1)',\ldots,(x_k^M)'\right)'$ se factorise souvent en

$$p(z_k|x_k) = \prod_{m=1}^{M} l_m(z_k|x_k^1,\ldots,x_k^m), \quad (2.43)$$

où les fonctions $l_m(z_k|.)$ permettent de définir des vraisemblances intermédiaires d'un sous-ensemble du vecteur d'état de plus en plus important au fur et à mesure que $m \longrightarrow M$.

Sous ces dernières hypothèses, les étapes élémentaires constitutives de l'algorithme de filtrage partitionné peuvent être transfomées en [MacCormick, 2000, §7.6, p. 131] :

1. Disposant du nuage pondéré $\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}$ représentant $p(x_{k-1}|z_{1:k-1})$, échantillonner de manière indépendante $x_k^{1(i)} \sim p(x_k^1|x_{k-1}^{1(i)})$, et associer à $(x_k^{1(i)}, x_{k-1}^{2(i)},\ldots,x_{k-1}^{M(i)})$ le poids $\tau_1^{(i)} \propto w_{k-1}^{(i)} l_1(z_k|x_k^{1(i)})$;
2. Rééchantillonner $\{(x_k^{1(i)}, x_{k-1}^{2(i)},\ldots,x_{k-1}^{M(i)}), \tau_1^{(i)}\}$ en l'ensemble de particules équipondérées $\{(\tilde{x}_k^{1(i)}, \tilde{x}_{k-1}^{2(i)},\ldots,\tilde{x}_{k-1}^{M(i)}), \tilde{\tau}_1^{(i)} = \frac{1}{N}\}$; renommer $\{(\tilde{x}_k^{1(i)}, \tilde{x}_{k-1}^{2(i)},\ldots,\tilde{x}_{k-1}^{M(i)}), \tilde{\tau}_1^{(i)} = \frac{1}{N}\}$ en $\{(x_k^{1(i)}, x_{k-1}^{2(i)},\ldots,x_{k-1}^{M(i)}), \tau_1^{(i)} = \frac{1}{N}\}$;
3. Échantillonner de manière indépendante $x_k^{2(i)} \sim p(x_k^2|x_{k-1}^{2(i)})$, et associer à $(x_k^{1(i)}, x_k^{2(i)}, x_{k-1}^{3(i)},\ldots,x_{k-1}^{M(i)})$ le poids $\tau_2^{(i)} \propto \tau_1^{(i)} l_2(z_k|x_k^{1(i)}, x_k^{2(i)})$;
4. Rééchantillonner [...] ; renommer [...] ;

$\vdots$

$(2M-1)$. À l'issue de l'échantillonnage de $x_k^{M(i)} \sim p(x_k^M|x_{k-1}^{M(i)})$ et de la mise à jour des poids $w_k^{(i)} \propto \tau_{M-1}^{(i)} l_M(z_k|x_k^{1(i)},\ldots,x_k^{M(i)})$, l'ensemble de particules pondérées $\{x_k^{(i)}, w_k^{(i)}\}$ obtenu est une représentation cohérente de $p(x_k|z_{1:k})$.

D-3 Échantillonnage hiérarchisé La stratégie d'échantillonnage hiérarchisé, développée dans [Pérez et al., 2004], peut être vue comme une généralisation du second algorithme d'échantillonnage partitionné présenté précédemment. La dynamique du système est supposée décomposable selon l'équation (2.38), sans qu'aucune restriction ne soit placée sur les fonctions $\tilde{d}_m(\cdot|\cdot)$, $m = 1, \ldots, M$. La fonction de mesure $p(z_k|x_k)$ est quant à elle supposée factorisable en un produit de M vraisemblances élémentaires $p_1(z_k|\cdot), \ldots, p_M(z_k|\cdot)$ – *e.g.* dans le cas où z_k est constitué de M informations sensorielles conditionnellement indépendantes étant donné x_k,

$$p(z_k|x_k) = \prod_{m=1}^{M} p_m(z_k|x_k) = \prod_{m=1}^{M} p_m(z_k^m|x_k) \tag{2.44}$$

–, telles que chaque vraisemblance $p_m(z_k|\cdot)$ puisse être incorporée après l'application de la dynamique $\tilde{d}_m(\cdot|\cdot)$. Une autre caractéristique essentielle de l'échantillonnage hiérarchisé est que les particules relatives aux vecteurs auxiliaires $\xi_1, \ldots, \xi_{M-1}$ et au vecteur d'état x_k ne sont pas échantillonnées selon les dynamiques $\tilde{d}_1(\cdot|\cdot), \ldots, \tilde{d}_{M-1}(\cdot|\cdot)$ et $\tilde{d}_M(\cdot|\cdot)$ mais selon des fonctions d'importance $\tilde{q}_1(\cdot|\cdot), \ldots, \tilde{q}_M(\cdot|\cdot)$ liées à la fonction d'importance $q(x_k|x_{k-1}, z_k)$ par

$$q(x_k|x_{k-1}, z_k) = \int \tilde{q}_1(\xi_1|x_{k-1}, z_k^1)\tilde{q}_2(\xi_2|\xi_1, z_k^2)\ldots\tilde{q}_M(x_k|\xi_{M-1}, z_k^M)d\xi_1 d\xi_2 \ldots d\xi_{M-1}. \tag{2.45}$$

L'algorithme s'écrit alors comme indiqué dans la Table 2.4.

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= HIERARCH $\left([\{x_{k-1}^{(i)}, w_{k-1}^{(i)}, \}]_{i=1}^N, z_k = ((z_k^1)', \ldots (z_k^M)')'\right)$

SI $k = 0$ **(INITIALISATION) ALORS**
  Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
FIN SI
Poser $\{\xi_0^{(i)}, \tau_0^{(i)}\} = \{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}$
SI $k \geq 1$ **ALORS**
  POUR $m = 1, \ldots, M$, **FAIRE**
    POUR $i = 1, \ldots, N$, **FAIRE**
      Échantillonner de manière indépendante $\xi_m^{(i)} \sim \tilde{q}_m(\xi_m|\xi_{m-1}^{(i)}, z_k^m)$
      Calculer les poids $\tau_m^{(i)} \propto \tau_{m-1}^{(i)} \frac{p_m(z_k^m|\xi_m^{(i)})\tilde{d}_m(\xi_m^{(i)}|\xi_{m-1}^{(i)})}{\tilde{q}_m(\xi_m^{(i)}|\xi_{m-1}^{(i)}, z_k^m)}$ préalablement à leur normalisation de telle sorte que $\sum_{i=1}^N \tau_m^{(i)} = 1$
    FIN POUR
    Réaffecter l'ensemble de particules pondérées $\{\xi_m^{(i)}, \tau_m^{(i)}\}$ avec l'ensemble de particules équipondérées équivalent obtenu par rééchantillonnage
  FIN POUR
  Fin : $\{x_k^{(i)}, w_k^{(i)}\} = \{\xi_M^{(i)}, \tau_M^{(i)}\}$
FIN SI

TAB. 2.4 – Filtre hiérarchisé pour la fusion de M modalités d'observation $z_1, \ldots, z_M$

E Réduction de variance par Rao-Blackwellisation

Lorsqu'une partie du vecteur d'état, conditionnellement à d'autres composantes, peut être traitée par une méthode optimale telle que le filtre de Kalman, il est opportun de ne pas résoudre le problème de filtrage en adoptant une solution « purement particulaire ». L'application de la méthode de Rao-Blackwellisation permet de tirer parti de cette propriété, et de réduire ainsi la variance du filtre [Casella et al., 1996, Doucet, 1998, Gustafsson et al., 2002].

Dans cette méthode, on suppose donc que le vecteur d'état x_k peut se diviser en deux parties u_k et v_k telles que

$$p(u_k|v_k, u_{k-1}, v_{k-1}) = p(u_k|u_{k-1}), \tag{2.46}$$

de sorte que l'équation de dynamique du système s'écrit

$$p(x_k|x_{k-1}) = p(u_k, v_k|u_{k-1}, v_{k-1}) = p(u_k|u_{k-1})p(v_k|u_{k-1}, v_{k-1}). \tag{2.47}$$

Si on suppose également que la distribution *a posteriori* conditionnelle $p(v_k|u_k, z_{1:k})$ peut être exprimée analytiquement, alors, en se basant sur la distribution objectif $p(x_k|z_{1:k})$ qui s'écrit

$$p(x_k|z_{1:k}) = p(u_k, v_k|z_{1:k}) = p(v_k|u_k, z_{1:k})p(u_k|z_{1:k}), \tag{2.48}$$

on peut simplement marginaliser sur v_k pour se focaliser sur l'estimation de $p(u_k|z_{1:k})$ dont l'espace est réduit.

Dans un premier temps, on construit une approximation particulaire de $p(u_k|z_{1:k})$,

$$p(u_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(u_k - u_k^{(i)}). \tag{2.49}$$

Ensuite, la densité marginale de $v_k|z_{1:k}$, qui s'écrit

$$p(v_k|z_{1:k}) \quad = \quad \int p(v_k|u_k, z_{1:k})p(u_k|z_{1:k})du_k, \tag{2.50}$$

est approchée, grâce à (2.49), par un mélange de lois calculables

$$\begin{aligned} p(v_k|z_{1:k}) &\simeq \int p(v_k|u_k, z_{1:k}) \sum_{i=1}^{N} w_k^{(i)} \delta(u_k - u_k^{(i)}) du_k \\ &\simeq \sum_{i=1}^{N} w_k^{(i)} p(v_k|u_k^{(i)}, z_{1:k}). \end{aligned} \tag{2.51}$$

Des études théoriques dans [Doucet et al., 2000] et [Doucet et al., 2001b] ont montré que la méthode de Rao-Blackwellisation apporte bien un gain sur l'efficacité du filtrage, en diminuant la variance par rapport à la variance obtenue par un estimateur standard non Rao-Blackwellisé.

F Compléments

De nombreuses autres stratégies ont été développées dans la littérature afin de gérer plus efficacement la population de particules, qui sortent du cadre de ce mémoire. Citons par exemple les filtres à mémoire limitée ou à oubli exponentiel, les méthodes de *prior editing / prior boosting* qui intègrent des tests d'acceptation des particules, ou l'ajout d'étapes de méthodes de Monte Carlo par Chaîne de Markov (MCMC) pour une meilleure diffusion des particules et une intégration graduelle de l'information apportée par la mesure dans la fonction d'importance.

2.2 Stratégies d'échantillonnage « simples »

Comme indiqué au §2.1.3–C, dès lors que la procédure de rééchantillonnage est sélectionnée dans l'algorithme générique SIR, tout filtre particulaire peut être dérivé par le choix d'une fonction d'importance dont le support inclut celui de la loi objectif. Ce choix a bien sûr des conséquences sur l'efficacité du filtre et la qualité de l'estimé. Dans cette section, nous présentons tout d'abord deux types de fonctions d'importance « simples », au sens où elle sont basées uniquement sur la dynamique ou sur les mesures. Des extensions de ces stratégies ainsi que des mécanismes permettant d'améliorer l'efficacité du filtrage sont également survolées.

2.2.1 Fonction d'importance basée sur la dynamique (FID)

Les premiers filtres particulaires proposés dans la littérature, tels le *filtre bootstrap* [Gordon et al., 1993] ou plus récemment la CONDENSATION – pour "Conditional Density Propagation" – [Isard et al., 1998a], peuvent être vus comme le cas particulier de l'algorithme SIR où la fonction d'importance est relative à la dynamique du système, *i.e.*

$$q(x_k|x_{k-1}^{(i)}, z_k) = p(x_k|x_{k-1}^{(i)}). \tag{2.52}$$

Les poids d'importance s'écrivent alors

$$w_k^{(i)} \quad \propto \quad w_{k-1}^{(i)} p(z_k|x_k^{(i)}). \tag{2.53}$$

Si de plus, un rééchantillonnage est implémenté à chaque instant, les poids des particules sont tous égaux à $\frac{1}{N}$ avant leur « propagation » par l'échantillonnage selon la dynamique, de sorte l'équation précédente se simplifie en

$$w_k^{(i)} \propto p(z_k|x_k^{(i)}). \tag{2.54}$$

Sur le plan algorithmique, cette stratégie présente l'intérêt de ne pas devoir mémoriser les poids d'importance d'un instant à l'autre.

L'algorithme de CONDENSATION ainsi obtenu (Table 2.5) adopte une structure prédiction / mise à jour comparable à celle du filtre de Kalman. En effet, la densité ponctuelle $\sum_{i=1}^{N} w_{k-1}^{(i)} \delta(x_k - x_k^{(i)})$ approxime la loi de prédiction $p(x_k|z_{1:k-1})$, et la mise à jour des poids selon (2.53) rappelle la formule de Bayes sous-jacente à l'étape de mise

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= CONDENSATION($[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}, \}]_{i=1}^N, z_k$)

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
3: **FIN SI**
4: **SI** $k \geq 1$ **ALORS**
5: **POUR** $i = 1, \ldots, N$, **FAIRE**
6: « Propager » la particule $x_{k-1}^{(i)}$ en simulant

$$x_k^{(i)} \sim p(x_k | x_{k-1}^{(i)})$$

7: Mettre à jour le poids $w_k^{(i)}$ selon l'équation

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(z_k | x_k^{(i)})$$

préalablement à une étape de normalisation assurant que $\sum_{i=1}^N w_k^{(i)} = 1$
8: **FIN POUR**
9: Le nuage $\{x_k^{(i)}, w_k^{(i)}\}_{i=1\ldots N}$ permet d'approcher la loi de filtrage par

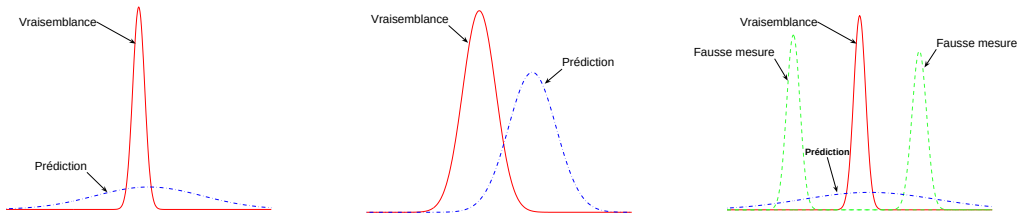
$$p(x_k | z_{1:k}) \simeq \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)})$$

10: De manière systématique où dès lors que $\frac{1}{\sum_{i=1}^N (w_k^{(i)})^2} < \textit{seuil}$, rééchantillonner $\{x_k^{(i)}, w_k^{(i)}\}$ selon $P(\tilde{x}_k^{(i)} = x_k^{(j)}) = w_k^{(j)}$, ce qui conduit à un ensemble de particules pondérées $\{\tilde{x}_k^{(i)}, \frac{1}{N}\}$ tel que $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ et $\frac{1}{N} \sum_{i=1}^N \delta(x_k - \tilde{x}_k^{(i)})$ approximent $p(x_k | z_{1:k})$; affecter $x_k^{(i)}$ et $w_k^{(i)}$ avec $\tilde{x}_k^{(i)}$ et $\frac{1}{N}$
11: **FIN SI**

TAB. 2.5 – Algorithme de CONDENSATION

à jour de l'estimé de Kalman. Bien que l'utilisation de la dynamique comme fonction d'importance présente l'avantage de simplifier la mise en œuvre de l'algorithme de filtrage, elle pose des problèmes significatifs d'efficacité et de précision de l'estimation. En effet, chaque particule $x_k^{(i)}$ est distribuée selon la prédiction de sa particule prédecesseur $x_{k-1}^{(i)}$ au moyen du modèle de dynamique $p(x_k | x_{k-1}^{(i)})$ sans que ne soit prise en compte l'observation courante z_k. Ceci entraîne un risque de mauvaise couverture des zones fortement vraisemblables vis à vis de la mesure, de sorte que l'espace d'état risque d'être mal exploré et la précision de l'estimation diminuée. À titre d'exemple, en cas de mauvais recouvrement entre la fonction d'importance et la fonction de mesure (Figure 2.1-(b)), typiquement lors de sauts dans la dynamique de la cible ou en sortie d'occultation, la non prise en compte de la nouvelle mesure dans la diffusion des particules, conduit à un nuage ne contenant aucune particule valide et par conséquent à une mauvaise représentation de la distribution. Pour une dynamique peu informative alors que les modes de la vraisemblance sont très prononcés (Figure 2.1-(a)), le rééchantillonnage systématique élimine une majorité des particules et duplique les quelques particules les plus vraisemblables. Cette situation conduit à une perte significative de diversité des particules et par conséquent à un appauvrissement important des états.

De plus, ce mode de diffusion (peu informatif) des particules, confère au filtre, une grande sensibilité aux fausses mesures (Figure 2.1-(c)). Il est important de noter que la vraisemblance qui est une loi de probabilité sur les mesures est vue ici comme une fonction de l'état. La normalisation de cette fonction dans les graphes de la Figure 2.1 n'est qu'une commodité graphique.



(a) (b) (c)

FIG. 2.1 – Fonction de vraisemblance et densité de prédiction pour diverses situations : (a) dynamique peu informative et observation étroite, (b) incohérence de l'observation vis à vis de la densité de prédiction et (c) dynamique trop peu informative en présence de fausses mesures

En conclusion, bien que très simple à mettre en œuvre et peu coûteuse en temps de calcul, la stratégie CONDENSATION présente le risque d'être peu efficace notamment pour un suivi visuel dans le contexte de la robotique. La propagation des particules « en aveugle » par rapport aux observations rend le filtre sensible aux fausses mesures et conduit à une dégénérescence du nuage de particules qui ne permet plus d'approximer correctement la distribution de filtrage. Pour ces raisons, nous considérons dans la section suivante une fonction d'importance basée sur les mesures.

2.2.2 Fonction d'importance basée sur les mesures (FIM)

Une alternative à la stratégie précédente consiste donc à définir

$$q(x_k|x_{k-1}^{(i)}, z_k) = q(x_k|z_k), \tag{2.55}$$

de sorte qu'à chaque instant k, les particules $x_k^{(i)}$ –ou bien seulement certaines de leur composantes– sont échantillonnées selon une information issue de l'observation. Dans le cas d'un suivi visuel de personne par exemple, il peut s'agir d'un détecteur de visage qui renseigne sur les positions potentielles de visages dans l'image. Le calcul des poids d'importance résultant de ce choix nécessite de pouvoir évaluer ponctuellement à une constante près la fonction d'importance choisie ainsi que d'évaluer la densité de probabilité de chaque particule conditionnée sur son passé. Il s'écrit

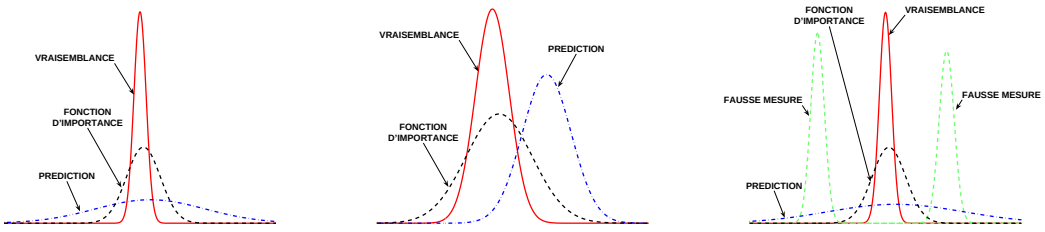
$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|z_k)}. \tag{2.56}$$

En suivi visuel, la première stratégie ayant pris en compte la mesure dans la fonction d'importance est l'ICONDENSATION [Isard et al., 1998b]. Dans la suite de ce document, nous désignerons par MSIR pour *Mesure SIR*, les algorithmes de filtrage particulaire obtenus à partir de l'algorithme SIR pour une fonction d'importance de la forme (2.55) et une mise à jour des poids d'importance au moyen de (2.56).

Dès lors que la fonction d'importance définie pour ces stratégies est plus informative que la fonction de prédiction (Figure 2.2-(a)), l'échantillonnage des particules conduit à un nuage plus concentré sur le mode de la fonction de vraisemblance et ainsi à une meilleure approximation de la distribution.

Dans le cas évoqué précédemment, où la fonction de prédiction et la fonction de mesure sont incohérentes (Figure 2.2-(b)), les particules mieux positionnées sur le pic de vraisemblance par une fonction d'importance relative à la mesure favorise une meilleure représentation de la densité *a posteriori* que dans le cas d'une diffusion des particules par la dynamique.

Enfin, sous l'hypothèse d'une fonction d'importance peu sensible aux fausses mesures (Figure 2.2-(c)), cette stratégie devient plus robuste qu'une stratégie utilisant la dynamique du système pour fonction d'importance.



(a) (b) (c)

FIG. 2.2 – Fonction de vraisemblance, fonction d'importance et densité de prédiction pour diverses situations : (a) dynamique peu informative et observation très fine, (b) fonction d'importance concordante avec l'observation mais incohérente vis à vis de la densité de prédiction et (c) dynamique peu informative en présence de fausses mesures avec fonction d'importance correcte

Cependant, une stratégie dont la fonction d'importance est uniquement basée sur des mesures soulève d'autres problèmes. En premier lieu, il est bien sûr indispensable de définir une telle fonction ; dans certains cas où on exploite une information intermittente issue de l'observation, il peut s'agir d'une contrainte difficile à satisfaire. De plus, des fausses mesures peuvent conduire à une fonction d'importance multi-modale qui risque alors de disperser inutilement un nuage de particules qui représentait correctement la distribution *a posteriori*. Enfin, quelle que soit la situation, rien n'empêche qu'une particule $x_k^{(i)}$ positionnée selon l'observation courante (même mono-modale) soit incompatible avec sa particule prédécesseur $x_{k-1}^{(i)}$ du point de vue de la dynamique du processus d'état. Du fait que $p(x_k^{(i)}|x_{k-1}^{(i)})$ prend de faibles valeurs, une telle particule

est alors faiblement pondérée dans (2.56) même si elle est fortement vraisemblable vis à vis de l'observation.

En conclusion, bien que le choix d'une fonction d'importance basée sur les mesures apporte généralement une meilleure précision qu'une fonction uniquement basée sur la dynamique, certaines difficultées apparaissent. Dans la section suivante nous présentons des stratégies pouvant être mises en œuvre pour pallier les difficultés rencontrées ainsi que des stratégies permettant d'améliorer l'efficacité du filtrage.

2.2.3 Extensions

Dans le but d'augmenter la qualité de l'estimation et d'améliorer l'efficacité du filtre, nous proposons quelques extensions des fonctions d'importance introduites dans les §2.2.1 et §2.2.2, puis présentons des stratégies essentiellement basées sur le rééchantillonnage.

A Combinaison de fonctions d'importance

Nous avons vu que l'utilisation de la dynamique du système comme fonction d'importance entraîne souvent une inefficacité du nuage de particules pour représenter la distribution *a posteriori*, notamment lorsque la fonction de mesure est fine et que la dynamique est peu informative. La stratégie opposée, qui consiste en la définition d'une fonction d'importance uniquement basée sur les mesures, apporte une amélioration dans l'approximation de la densité *a posteriori* en regroupant le nuage de particules dans les zones de forte vraisemblance vis à vis de la mesure. En pratique, et notamment pour un suivi visuel, les fonctions d'importance dérivent souvent d'un processus de détection imparfait, pouvant parfois ommettre certains pics de vraisemblance ou bien être sensible à des pics issus de fausses mesures. Dans ce cas, la fonction d'importance, qui ne peut pas être obtenue à chaque instant ou qui prend une forme multi-modale, a tendance à disperser les particules dans l'espace d'état, conduisant à une mauvaise représentation de la distribution $p(x_k|z_{1:k})$. Une solution simple à ces problèmes est de définir la fonction d'importance comme un mélange d'une fonction d'importance basée sur les mesures et d'une fonction d'importance basée sur la dynamique ([Isard et al., 1998b], [Pérez et al., 2004]) :

$$q(x_k|x_{k-1}^{(i)}, z_k) = (1-\alpha)p(x_k|x_{k-1}^{(i)}) + \alpha q(x_k|z_k). \tag{2.57}$$

Ainsi, un pourcentage α de particules sont échantillonnées selon l'observation courante alors que les particules restantes suivent la dynamique du système. De plus, en cas de fausses mesures ou d'absence de mesure lors de la définition de la fonction d'importance, une partie du nuage de particules continue à évoluer selon la dynamique du système permettant alors de conserver une meilleure représentation de la distribution *a posteriori*.

De la même manière, il peut être intéressant d'échantillonner une partie des particules selon une connaissance *a priori* $q_0(x_k)$ de façon qu'elles soient positionnées indépendamment de l'état précédent et des mesures ([Isard et al., 1998b],

[Pérez et al., 2004]) :

$$q(x_k|x_{k-1}^{(i)}) = (1-\beta)p(x_k|x_{k-1}^{(i)}) + \beta q_0(x_k) \tag{2.58}$$

En choisissant pour $q_0(x_k)$ la fonction d'initialisation *a priori* $q_0(x_k) = p(x_0)$ du filtre, cette combinaison peut permettre une réinitialisation automatique en cas de perte de la cible suite à d'importants échecs de mesure ou à un mouvement de la cible pendant une occultation. Lorsqu'on ne dispose pas d'une fonction $q_0(x_k)$ pertinente, on peut recourir à une fonction d'importance $q(x_k|z_k)$ permettant de positionner certaines particules dans des zones vraisemblables de l'espace sans tenir compte de la dynamique, et ainsi donner au filtre la possibilité de retrouver la cible en cas de perte [Isard et al., 1998b]. Comme dans [Pérez et al., 2004], on peut aussi choisir une loi uniforme (*flat prior*) $q_0(x_k) = \mathcal{U}(x_k)$. Dans le cas d'un espace fini et borné, les rares mouvements erratiques de la cible peuvent ainsi être capturés.

Les stratégies d'échantillonnage (2.57)–(2.58) peuvent être mises en œuvre indépendamment ou être combinées en

$$q(x_k|x_{k-1}^{(i)}, z_k) = \alpha q(x_k|z_k) + \beta q_0(x_k) + (1-\alpha-\beta)p(x_k|x_{k-1}^{(i)}). \tag{2.59}$$

Le positionnement des particules dans l'espace d'état peut être amélioré au moyen d'une combinaison adéquate de fonctions d'importance, mais l'efficacité d'un filtre peut encore être augmentée en utilisant des rééchantillonnages de manière apropriée.

B Améliorations par le rééchantillonnage

Comme indiqué précédemment, le rééchantillonnage permet la redistribution d'un nuage de particules guidée par une fonction ou un vecteur de poids, dans le but d'obtenir une représentation plus fidèle d'une loi. Classiquement, on trouve un rééchantillonnage (conditionné sur l'évaluation de $N_{eff} < seuil$) à la fin de l'algorithme générique de filtrage SIR pour empêcher la dégénérescence du nuage de particules provoquée par l'échantillonnage pondéré séquentiel. Il peut cependant être utilisé à d'autres niveaux du filtre. Dans la méthode d'échantillonnage hiérarchisé utilisée par [Pérez et al., 2004] par exemple, un « rééchantillonnage systématique » [5] est placé entre les étapes de simulation des diverses composantes du vecteur d'état. Ainsi, les particules sont redistibuées selon les zones vraisemblables de la partition courante de l'espace d'état, ce qui conduit à un nuage de particules plus adapté au traitement de la partition suivante. De la même façon, l'échantillonnage partitionné [MacCormick et al., 2000] inclut un *rééchantillonnage pondéré* entre chaque étape, de façon à repositionner les particules selon une fonction qui rend compte de la vraisemblance de la partition courante de l'état vis à vis de l'observation.

Malgré le gain en efficacité apporté par l'utilisation de ces stratégies, les filtres de type MSIR souffrent d'un problème majeur lié à la définition de la fonction d'importance. En effet, comme indiqué au §2.2.2, le fait qu'ils positionnent tout ou partie des

[5] Nous rappelons que le terme « systématique » se rapporte ici à la méthode de rééchantillonnage [Kitagawa, 1996], et non au fait que ce rééchantillonnage soit appliqué à chaque instant.

composantes des particules seulement à partir de l'observation peut conduire à une incompatibilité de ces particules avec leurs particules prédécesseurs du point de vue de la dynamique du système. Une alternative intéressante proposée dans [Torma et al., 2003] permet de résoudre cette incohérence en mettant en œuvre des mécanismes basés comme précédemment sur la combinaison du partitionnement de l'espace d'état avec des ré-échantillonnages.

Les algorithmes proposés dans [Torma et al., 2003] permettent de prendre en compte des modèles dynamiques d'ordre supérieur ou égal à 2, où un sous-vecteur u_k de x_k – appelé « partie innovation[6] » – obéit à une équation d'état stochastique en x_{k-1} alors que le sous-vecteur complémentaire de u_k – qualifié de « partie historique » – est une fonction déterministe de x_{k-1}. En d'autres termes, l'évolution de $x_k = (u'_k, h'_k)'$ est définie par :

$$\begin{aligned} u_k &= f_1(x_{k-1}) + s_k \\ h_k &= f_2(x_{k-1}) \end{aligned} \tag{2.60}$$

où s_k désigne le bruit de dynamique.

Les auteurs supposent que la fonction d'importance est réduite à la partie innovation, de la forme $q(u_k|z_k)$, et que la fonction de vraisemblance satisfait $p(z_k|x_k) = p(z_k|u_k)$. Ce contexte est particulièrement bien adapté au suivi visuel, car les représentations d'état de modèles de dynamique linéaires auto-régressifs sont semblables à (2.50) et car la « partie historique » n'intervient pas dans le lien état-mesure.

Plusieurs algorithmes sont proposés de façon à éviter toute contradiction entre $u_k^{(i)} \sim q(u_k|z_k)$ et le passé $x_{k-1}^{(i)}$. Le premier est une extension de l'algorithme MSIR intégrant un rééchantillonnage des « parties innovations » guidé par leurs probabilités d'occurence, suivi de l'échantillonnage d'un passé – et donc d'une « partie historique » – plausible. Cet algorithme, nommé HSSIR – pour *History Sampling SIR* –, est présenté Table 2.6.

Pour le justifier, supposons que l'ensemble de particules pondérées relatives à la densité *a posteriori* à l'instant $k-1$ s'écrive $\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}$, et désignons par $\{x_k^{(m,n)}, w_k^{(m,n)}\}$ l'ensemble des particules représentant $p(x_k|z_{1:k})$, avec

$$x_k^{(m,n)} = \begin{pmatrix} u_k^{(n)} \\ h_k^{(m)} \end{pmatrix} \sim q_x(x_k|x_{k-1}^{(m)}, z_k) \;\Leftrightarrow\; \begin{cases} u_k^{(n)} \sim q(u_k|z_k) \\ h_k^{(m)} = f_2(x_{k-1}^{(m)}) \end{cases} \tag{2.61}$$

et

$$w_k^{(m,n)} \propto w_{k-1}^{(m)} \frac{p(z_k|x_k^{(m,n)}) p(x_k^{(m,n)}|x_{k-1}^{(m)})}{q_x(x_k^{(m,n)}|x_{k-1}^{(m)}, z_k)}. \tag{2.62}$$

Du fait que $p(z_k|x_k^{(m,n)}) = p(z_k|u_k^{(n)})$ et que $\frac{p(x_k^{(m,n)}|x_{k-1}^{(m)})}{q_x(x_k^{(m,n)}|x_{k-1}^{(m)}, z_k)} = \frac{p(u_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)}|z_k)}$, il vient

$$w_k^{(m,n)} \propto w_{k-1}^{(m)} \frac{p(z_k|u_k^{(n)}) p(u_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)}|z_k)}. \tag{2.63}$$

[6] Le sens donné par les auteurs au vocable « innovation » doit être distingué du terme français relatif à l'erreur de prédiction sur la mesure, dont la traduction anglo-saxonne serait "residual".

En se basant sur cette formulation, l'algorithme HSSIR génère l'ensemble des particules représentant $p(x_k|z_{1:k})$ en deux temps. Tout d'abord, les indices n des innovations les plus probables sont sélectionnés, compte tenu des vraisemblances de celles-ci par rapport à l'observation courante ainsi que de la densité de prédiction approximée à partir du nuage à l'instant précédent. Puis, pour chaque innovation, un passé plausible d'indice m est échantillonné parmi tous les historiques possibles. Ainsi, la cohérence entre les parties « innovation » et « historique » des particules est conservée malgré l'échantillonnage des innovations selon une fonction d'importance uniquement basée sur les mesures. Mathématiquement, les fonctions qui guident ces rééchantillonnages peuvent être déterminées en notant que l'ensemble $\{x_k^{(m,n)}, w_k^{(m,n)}\}$ de N^2 particules est transformé en l'ensemble équipondéré $\{x_k^{(I_k^{(i)},J_k^{(i)})}, \frac{1}{N}\}$ de N particules si et seulement si $\mathbb{P}(I_k^{(i)} = m, J_k^{(i)} = n) = w_k^{(m,n)}$. Or, $\mathbb{P}(I_k^{(i)} = m, J_k^{(i)} = n)$ est égal au produit $\mathbb{P}(I_k^{(i)} = m|J_k^{(i)} = n)\,\mathbb{P}(J_k^{(i)} = n)$, où la probabilité $\mathbb{P}(J_k^{(i)} = n)$ d'occurence de l'innovation $u_k^{(n)}$ est obtenue par la marginalisation

$$\mathbb{P}(J_k^{(i)} = n) = \sum_{m=1}^{N} w_k^{(m,n)} \tag{2.64}$$

$$\propto \sum_{m=1}^{N} w_{k-1}^{(m)} \frac{p(z_k|u_k^{(n)})p(u_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)}|z_k)}, \tag{2.65}$$

$$\text{soit } \mathbb{P}(J_k^{(i)} = n) = \frac{\frac{p(z_k|u_k^{(n)})}{q(u_k^{(n)}|z_k)} \sum_{m=1}^{N} w_{k-1}^{(m)} p(u_k^{(n)}|x_{k-1}^{(m)})}{\sum_{l=1}^{N} \frac{p(z_k|u_k^{(l)})}{q(u_k^{(l)}|z_k)} \sum_{m=1}^{N} w_{k-1}^{(m)} p(u_k^{(l)}|x_{k-1}^{(m)})}. \tag{2.66}$$

Il vient alors

$$\mathbb{P}(I_k^{(i)} = m|J_k^{(i)} = n) = \frac{w_k^{(m,n)}}{\mathbb{P}(J_k^{(i)} = n)} \tag{2.67}$$

$$\propto w_{k-1}^{(m)} p(u_k^{(n)}|x_{k-1}^{(m)}), \tag{2.68}$$

$$\text{soit } \mathbb{P}(I_k^{(i)} = m|J_k^{(i)} = n) = \frac{w_{k-1}^{(m)} p(u_k^{(n)}|x_{k-1}^{(m)})}{\sum_{r=1}^{N} w_{k-1}^{(r)} p(u_k^{(n)}|x_{k-1}^{(r)})}. \tag{2.69}$$

Les expressions (2.66) et (2.69) suggèrent donc de rééchantillonner $\{x_k^{(m,n)}, w_k^{(m,n)}\}$ en $\{x_k^{(I_k^{(i)},J_k^{(i)})}, \frac{1}{N}\}$ en suivant une procédure partitionnée telle que celle évoquée au §2.1.3–D-2. Elles apparaissent ainsi respectivement dans les items 7 et 8 de l'algorithme HSSIR Table 2.6.

En sélectionnant les paires d'innovations et d'historiques qui ont de fortes probabilitées de co-occurence, l'algorithme HSSIR permet généralement de réduire la variance de l'estimateur de la densité *a posteriori*. A l'inverse, l'introduction de rééchantillonnages dans un algorithme, produit une augmentation la variance, cf. §2.1.3–C. Afin

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= HSSIR($[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}, \}]_{i=1}^N, z_k$) avec $x_k = (u_k, v_k)$

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
3: **FIN SI**
4: **SI** $k \geq 1$ **ALORS**
5: Échantillonner $u_k^{(1)}, \ldots, u_k^{(i)}, \ldots, u_k^{(N)}$ i.i.d. selon $q(u_k|z_k)$.
6: **POUR** $i = 1, \ldots, N$, **FAIRE**
7: Échantillonner dans $(1 \ldots N)$ l'indice $J_k^{(i)}$ de la partie innovation u_k selon des poids proportionnels à

$$\left[\frac{p(z_k|u_k^{(1)})}{q(u_k^{(1)}|z_k)} \sum_{j=1}^N w_{k-1}^{(j)} p(u_k^{(1)}|x_{k-1}^{(j)}), \ldots, \frac{p(z_k|u_k^{(.)})}{q(u_k^{(.)}|z_k)} \sum_{j=1}^N w_{k-1}^{(j)} p(u_k^{(.)}|x_{k-1}^{(j)}), \ldots, \frac{p(z_k|u_k^{(N)})}{q(u_k^{(N)}|z_k)} \sum_{j=1}^N w_{k-1}^{(j)} p(u_k^{(N)}|x_{k-1}^{(j)})\right]$$

8: Échantillonner dans $(1 \ldots N)$ l'indice $I_k^{(i)}$ de la particule prédécesseur de $u_k^{J_k^{(i)}}$ selon des poids proportionnels à

$$\left[w_{k-1}^{(1)} p(u_k^{(J_k^{(i)})}|x_{k-1}^{(1)}), \ldots, w_{k-1}^{(.)} p(u_k^{(J_k^{(i)})}|x_{k-1}^{(.)}), \ldots, w_{k-1}^{(N)} p(u_k^{(J_k^{(i)})}|x_{k-1}^{(N)})\right]$$

9: Construire la particule en sélectionnant l'innovation et l'historique d'indices respectifs $J_k^{(i)}$ et $I_k^{(i)}$

$$x_k^{(i)} = \begin{pmatrix} u_k^{(J_k^{(i)})} \\ f_2(x_{k-1}^{(I_k^{(i)})}) \end{pmatrix}$$

10: **FIN POUR**
11: Le nuage $\{x_k^{(i)}, 1/N\}_{i=1\ldots N}$ permet d'approcher la loi de filtrage par

$$p(x_k|z_{1:k}) \simeq \frac{1}{N} \sum_{i=1}^N \delta(x_k - x_k^{(i)})$$

12: **FIN SI**

TAB. 2.6 – Algorithme de filtrage avec échantillonnage de l'historique (HSSIR)

de limiter ce problème et dans le but de diminuer encore la variance de l'estimateur, [Torma et al., 2003] propose également une version "Rao-Blackwellisée" de l'algorithme HSSIR. Cet algorithme que nous désignerons par RBHSSIR – pour *Rao-Blackwellised History Sampling SIR* – est une adaptation du HSSIR permettant de supprimer le ré-échantillonnage des « parties innovations ». Comme indiqué dans la Table 2.7, après avoir échantillonné les innovations selon la fonction d'importance (item 6), seul un passé plausible pour chaque innovation est rééchantillonné (item 7). Afin de conserver une approximation cohérente de la densité *a posteriori*, chaque particule ainsi obtenue est affectée d'un poids calculé selon (2.66), cf. l'item 9 de l'algorithmes RBHSSIR (Table 2.7).

Enfin, une autre variante intéressante est proposée dans le cas de fonctions d'importance ne permettant d'échantillonner qu'une partie de l'innovation. En suivi visuel, par exemple, l'état de la cible est parfois constitué des valeurs présentes et passées

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= RBHSSIR($[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}, \}]_{i=1}^N, z_k$) avec $x_k = (u_k, v_k)$

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
3: **FIN SI**
4: **SI** $k \geq 1$ **ALORS**
5: **POUR** $i = 1, \ldots, N$, **FAIRE**
6: Échantillonner de manière indépendante $u_k^{(i)} \sim q(u_k|z_k)$
7: Échantillonner dans $(1 \ldots N)$ l'indice $I_k^{(i)}$ de la particule prédécesseur de $u_k^{(i)}$ selon les poids proportionnels à

$$\left[w_{k-1}^{(1)} p(u_k^{(i)}|x_{k-1}^{(1)}), \ldots, w_{k-1}^{(.)} p(u_k^{(i)}|x_{k-1}^{(.)}), \ldots, w_{k-1}^{(N)} p(u_k^{(i)}|x_{k-1}^{(N)})\right]$$

8: Construire la particule en concaténant à $u_k^{(i)}$ l'historique indicé par $I_k^{(i)}$

$$x_k^{(i)} = \begin{pmatrix} u_k^{(i)} \\ f_2(x_{k-1}^{(I_k^{(i)})}) \end{pmatrix}$$

9: Mettre à jour les poids, préalablement à leur normalisation, en posant

$$w_k^{(i)} \propto \frac{p(z_k|u_k^{(i)})}{q(u_k^{(i)}|z_k)} \sum_{l=1}^{N} w_{k-1}^{(l)} p(u_k^{(i)}|x_{k-1}^{(l)})$$

de sorte que $p(x_k|z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$
10: **FIN POUR**
11: **FIN SI**

TAB. 2.7 – Algorithme de filtrage Rao-Blackwellisé avec échantillonnage de l'historique (RBHSSIR)

d'un sous-vecteur relatif aux paramètres de translation et d'une partie complémentaire relative aux paramètres de déformation (orientation, échelle, ...) Souvent, la fonction d'importance mise en œuvre ne renseigne que sur la composante translation (détection de blobs couleur, détection de visages, ...). Cette extension de l'algorithme RBHSSIR nommée RBSSHSSIR –pour *Rao-Blackwellised Subspace History Sampling SIR*– permet de prendre en compte des vecteurs d'état de la forme :

$$x_k = \begin{pmatrix} u_k \\ v_k \\ h_k \end{pmatrix}, \; h_k = f_2(x_{k-1}), \tag{2.70}$$

où u_k correspond à la partie de l'innovation échantillonnée selon une fonction d'importance $q(u_k|z_k)$ et v_k est le sous-vecteur complémentaire de u_k dans l'innovation. Cette variante de l'algorithme RBHSSIR est obtenue en notant que la fonction d'importance relative à l'ensemble de l'innovation peut s'écrire

$$q(u_k, v_k|x_{k-1}^{(i)}, z_k) = q(u_k|x_{k-1}^{(i)}, z_k) q(v_k|u_k, x_{k-1}^{(i)}, z_k), \tag{2.71}$$

soit, d'après l'hypothèse concernant l'échantillonnage de u_k,

$$q(u_k, v_k|x_{k-1}^{(i)}, z_k) = q(u_k|z_k) q(v_k|u_k, x_{k-1}^{(i)}, z_k). \tag{2.72}$$

La pondération (2.62) associée à $x_k^{(m,n)}$ devient, en utilisant le fait que $\frac{p(x_k^{(m,n)}|x_{k-1}^{(m)})}{q_x(x_k^{(m,n)}|x_{k-1}^{(m)},z_k)} = \frac{p(u_k^{(n)},v_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)},v_k^{(n)}|x_{k-1}^{(m)},z_k)}$,

$$w_k^{(m,n)} \propto w_{k-1}^{(m)} \frac{p(z_k|u_k^{(n)},v_k^{(n)})p(u_k^{(n)},v_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)}|z_k)q(v_k^{(n)}|u_k^{(n)},x_{k-1}^{(m)},z_k)} \tag{2.73}$$

$$\propto w_{k-1}^{(m)} \frac{p(z_k|u_k^{(n)},v_k^{(n)})p(v_k^{(n)}|u_k^{(n)},x_{k-1}^{(m)})p(u_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)}|z_k)q(v_k^{(n)}|u_k^{(n)},x_{k-1}^{(m)},z_k)} \tag{2.74}$$

et se simplifie, dans la veine de (2.63), en

$$w_k^{(m,n)} \propto w_{k-1}^{(m)} \frac{p(z_k|u_k^{(n)},v_k^{(n)})p(u_k^{(n)}|x_{k-1}^{(m)})}{q(u_k^{(n)}|z_k)} \tag{2.75}$$

dès lors que $v_k^{(n)}$ est échantillonné de manière indépendante selon $v_k^{(n)} \sim q(v_k|u_k^{(n)},x_{k-1}^{(m)},z_k) = p(v_k|u_k^{(n)},x_{k-1}^{(m)})$.

L'algorithme RBSSHSSIR est présenté Table 2.8. Il convient de remarquer que l'échantillonnage de la composante v_k est situé après la sélection d'un passé vraisemblable vis à vis de la partie u_k de l'innovation. Dans le cas où u_k et v_k sont indépendants, le tir de v_k se limite à la propagation de v_{k-1} selon la dynamique $v_k^{(i)} \sim p(v_k|x_{k-1}^{(I_k^{(i)})})$. On montre que le RBSSHSSIR demeure valide pour des dynamiques du premier ordre, auquel cas il suffit de supprimer la partie $f_2(x_{k-1}^{(i)})$ de x_k, cf. la fin du §2.3.1.

En conclusion, les versions Rao-Blackwellisées RBHSSIR et RBSSHSSIR assurent une gestion plus efficace du nuage de particules, et permettent ainsi d'obtenir un estimateur de la densité *a posteriori* de meilleure qualité. Leur particularité est de rééchantillonner, pour chaque « innovation » sélectionnée selon la fonction d'importance, un passé vraisemblable du point de vue de la dynamique et d'un poids conséquent. Elles diffèrent de la ICONDENSATION par ce rééchantillonnage (item 7 dans les Tables 2.7 et 2.8), qui pourtant est nécessaire sous peine que l'ensemble de particules pondérées $\{x_k^{(i)}, w_k^{(i)}\}$ ne constitue pas une approximation cohérente de la densité *a posteriori* $p(x_k|z_{1:k})$. En outre, l'utilisation de la dynamique du système dans le rééchantillonnage rend inutile le recours à des fonctions d'importance combinant également la distribution *a priori* du vecteur d'état à l'instant initial.

2.3 Vers le cas optimal

Parallèlement aux stratégies proposées au §2.2.3, l'efficacité du filtrage peut être significativement améliorée en définissant une fonction d'importance approchant la fonction d'importance optimale (2.30) définie au §2.1.3–B.

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= RBSSHSSIR($[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}, \}]_{i=1}^N, z_k$)

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
3: **FIN SI**
4: **SI** $k \geq 1$ **ALORS**
5: **POUR** $i = 1, \ldots, N$, **FAIRE**
6: Échantillonner de manière indépendante $u_k^{(i)} \sim q(u_k|z_k)$.
7: Échantillonner dans $(1 \ldots N)$ l'indice $I_k^{(i)}$ de la particule prédécesseur de $u_k^{(i)}$ selon les poids proportionnels à

$$\left[w_{k-1}^{(1)} p(u_k^{(i)}|x_{k-1}^{(1)}), \ldots, w_{k-1}^{(.)} p(u_k^{(i)}|x_{k-1}^{(.)}), \ldots, w_{k-1}^{(N)} p(u_k^{(i)}|x_{k-1}^{(N)})\right]$$

8: Échantillonner de manière indépendante $v_k^{(i)} \sim p(v_k|u_k^{(i)}, x_{k-1}^{(I_k(i))})$
9: Construire la particule en concaténant $u_k^{(i)}$ et $v_k^{(i)}$ dans l'innovation et en sélectionnant l'historique d'indice $I_k^{(i)}$

$$x_k^{(i)} = \begin{pmatrix} u_k^{(i)} \\ v_k^{(i)} \\ f_2(x_{k-1}^{(I_k^{(i)})}) \end{pmatrix}$$

10: Mettre à jour les poids, préalablement à leur normalisation, en posant

$$w_k^{(i)} \propto \frac{p(z_k|u_k^{(i)}, v_k^{(i)})}{q(u_k^{(i)}|z_k)} \sum_{l=1}^{N} w_{k-1}^{(l)} p(u_k^{(i)}|x_{k-1}^{(l)})$$

de sorte que $p(x_k|z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$
11: **FIN POUR**
12: **FIN SI**

TAB. 2.8 – Algorithme de filtrage à sous espace Rao-Blackwellisé avec échantillonnage de l'historique (RBSSHSSIR)

2.3.1 Stratégie "Auxiliary"

Le §2.1.3–B mentionne une propriété essentielle de la stratégie récursive optimale, selon laquelle le poids $w_k^{*(i)}$ associé à chaque particule $x_k^{(i)}$ via (2.31) ne dépend que de la particule prédécesseur $x_{k-1}^{(i)}$. Il devient alors possible de calculer les poids $w_k^{*(i)}$ avant même la « propagation » des particules $x_{k-1}^{(i)}$ selon la fonction d'importance (2.30). Si besoin, l'efficacité globale de l'algorithme peut être augmentée en introduisant un rééchantillonnage auxiliaire de l'ensemble $\{x_{k-1}^{(i)}, w_k^{*(i)}\}$ – qui, accessoirement, représente le lisseur $p(x_{k-1}|z_{1:k})$ – de façon que les particules soient équipondérées préalablement à leur propagation par (2.30).

Le filtre à « particules auxiliaires » – APF = *Auxiliary Particle Filter* – introduit par Pitt et Shephard [Pitt et al., 1999] contemporainement à l'apparition de la ICONDENSATION, tire parti de cette propriété. Disposant d'une approximation $\hat{p}(z_k|x_{k-1}^{(i)})$ de $p(z_k|x_{k-1}^{(i)})$, laquelle repose donc sur la donnée de l'observation courante z_k, une « pondération auxiliaire » $\lambda_k^{(i)} \propto w_{k-1}^{(i)} \hat{p}(z_k|x_{k-1}^{(i)})$ – mimant le rôle de $w_k^{*(i)}$ – est

associée à chaque particule $x_{k-1}^{(i)}$. L'ensemble $\{x_{k-1}^{(i)}, \lambda_k^{(i)}\}$ est alors rééchantillonné en un nuage équipondéré $\{x_{k-1}^{(s^{(i)})}, \frac{1}{N}\}$, lequel est ensuite « propagé »jusqu'à l'instant k au moyen d'une fonction d'importance $\pi(x_k|x_{k-1}^{(s^{(i)})}, z_k)$. Contrairement au cas optimal, les pondérations $w_k^{(i)}$ des particules $x_k^{(i)}$ résultantes doivent être corrigées *a posteriori* de façon à prendre en compte la « distance » entre $\lambda_k^{(i)}$ et $w_k^{*(i)}$, ainsi que le fait que la fonction d'importance $\pi(x_k|x_{k-1}, z_k)$ retenue diffère de la fonction d'importance optimale $p(x_k|x_{k-1}, z_k)$.

Une interprétation immédiate en terme d'instanciation de l'algorithme générique SIR consiste à remarquer que l'ensemble des particules pondérées $\{(x_k^{(i)}, s^{(i)}), w_k^{(i)}\}$, où la variable auxiliaire $s^{(i)}$ désigne l'indice de la particule prédécesseur de $x_k^{(i)}$ à l'instant $k-1$, représente la densité conjointe $p(x_k, s|z_{1:k})$. En effet, par marginalisation selon s, il vient immédiatement l'approximation de la loi de filtrage $p(x_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)})$.

La règle de Bayes permet d'établir que

$$\begin{aligned} p(x_k, s|z_{1:k}) &\propto p(z_k|x_k)p(x_k, s|z_{1:k-1}) \\ &\propto p(z_k|x_k)p(x_k|s, z_{1:k-1})p(s|z_{1:k-1}) \\ &\propto p(z_k|x_k)p(x_k|x_{k-1}^{(s)})w_{k-1}^{(s)}. \end{aligned} \tag{2.76}$$

D'autre part, la fonction d'importance $q(x_k, s|z_{1:k})$ selon laquelle sont échantillonnés les couples $\{(x_k^{(i)}, s^{(i)})\}$ s'écrit

$$q(x_k, s|z_{1:k}) = q(x_k|s, z_{1:k})q(s|z_{1:k}) \tag{2.77}$$

avec

$$q(s|z_{1:k}) = \lambda_k^{(s)} \propto w_{k-1}^{(s)} \hat{p}(z_k|x_{k-1}^{(s)}) \tag{2.78}$$

– réalisé grâce au rééchantillonnage – et

$$q(x_k|s, z_{1:k}) = \pi(x_k|x_{k-1}^{(s)}, z_k). \tag{2.79}$$

Dès lors, le poids d'importance $w_k^{(i)}$ associé à chaque couple $(x_k^{(i)}, s^{(i)})$ devient

$$w_k^{(i)} \propto \frac{p(x_k^{(i)}, s^{(i)}|z_{1:k})}{q(x_k^{(i)}, s^{(i)}|z_{1:k})} \propto \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(s^{(i)})})}{\hat{p}(z_k|x_{k-1}^{(s^{(i)})})\pi(x_k^{(i)}|x_{k-1}^{(s^{(i)})}, z_k)}. \tag{2.80}$$

Si on suppose, comme dans l'algorithme original de Pitt et Shephard, que

$$\pi(x_k|x_{k-1}^{(s)}, z_k) = p(x_k|x_{k-1}^{(s)}) \tag{2.81}$$

– *i.e.* après le rééchantillonnage auxiliaire, les particules sont propagées selon la dynamique du système –, et que

$$\hat{p}(z_k|x_{k-1}^{(s)}) = p(z_k|\mu_k^{(s)}) \tag{2.82}$$

où $\mu_k^{(s)}$ caractérise la distribution de x_k conditionnée sur $x_{k-1}^{(s)}$, *e.g.* $\mu_k^{(s)}$ désigne l'espérance $\mu_k^{(s)} = \mathbb{E}_{p(.|x_{k-1}^{(s)})}(x_k)$ ou bien une particule $\mu_k^{(s)} \sim p(x_k|x_{k-1}^{(s)})$, alors il vient

$$w_k^{(i)} \propto \frac{p(z_k|x_k^{(i)})}{p(z_k|\mu_k^{(s^{(i)})})}. \tag{2.83}$$

Sous ces hypothèses, l'algorithme AUXILIARY est résumé dans la table 2.9.

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= AUXILIARY($[\{x_{k-1}^{(i)}, w_{k-1}^{(i)},\}]_{i=1}^N, z_k$)

SI $k = 0$ **(INITIALISATION) ALORS**
  Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
FIN SI
SI $k \geq 1$ **ALORS**
  POUR $i = 1, \ldots, N$, **FAIRE**
    À partir de l'approximation $\hat{p}(z_k|x_{k-1}^{(i)})$ – *e.g.* $\hat{p}(z_k|x_{k-1}^{(i)}) = p(z_k|\mu_k^{(i)})$, où $\mu_k^{(i)} \sim p(x_k|x_{k-1}^{(i)})$ ou bien $\mu_k^{(i)} = \mathbb{E}_{p(.|x_{k-1}^{(i)})}(x_k)$–, calculer les pondérations auxiliaires

$$\lambda_k^{(i)} \propto w_{k-1}^{(i)} \hat{p}(z_k|x_{k-1}^{(i)})$$

  FIN POUR
  POUR $i = 1, \ldots, N$, **FAIRE**
    Rééchantillonner $\{x_{k-1}^{(i)}, \lambda_k^{(i)}\}$ –ou, de manière équivalente, échantillonner les indices j des particules à l'instant $k-1$ selon $\mathbb{P}(s^{(i)} = j) = q(j|z_{1:k}) \triangleq \lambda_k^{(j)}$– de façon à obtenir l'ensemble équivalent de particules équipondérées $\{x_{k-1}^{s^{(i)}}, \frac{1}{N}\}$; $\sum_{i=1}^N \lambda_k^{(i)} \delta(x_{k-1} - x_{k-1}^{(i)})$ et $\frac{1}{N}\sum_{i=1}^N \delta(x_{k-1} - x_{k-1}^{(s^{(i)})})$ représentent $p(x_{k-1}|z_{1:k})$
  FIN POUR
  POUR $i = 1, \ldots, N$, **FAIRE**
    « Propager » les particules en échantillonnant de manière indépendante $x_k^{(i)} \sim p(x_k|x_{k-1}^{(s^{(i)})})$
    Mettre à jour les poids, préalablement à leur normalisation, en posant

$$w_k^{(i)} \propto \frac{p(z_k|x_k^{(i)})}{\hat{p}(z_k|x_{k-1}^{(s^{(i)})})} = \frac{p(z_k|x_k^{(i)})}{p(z_k|\mu_k^{(s^{(i)})})},$$

    de sorte que $p(x_k|z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$
  FIN POUR
FIN SI

TAB. 2.9 – Algorithme de filtrage à « particules auxiliaires » (AUXILIARY)

L'utilisation de l'AUXILIARY permet de mieux orienter les particules vers les zones pertinentes de l'espace d'état et ainsi d'obtenir une approximation plus efficace de la densité *a posteriori*. Cependant, lorsque les pics de vraisemblance coïncident avec le mode de la densité de prédiction, cet algorithme n'apporte pas de gain par rapport à une méthode comme la CONDENSATION. Au contraire, il est plutôt pénalisé par l'introduction du rééchantillonnage auxiliaire qui augmente la variance de l'estimateur. De plus, l'utilisation de la densité de prédiction pour échantillonner les particules peut

poser problème, notamment lorsque le modèle de dynamique utilisé est peu informatif par rapport à la vraisemblance. En effet, $\mu_k^{(s)}$ ne permet pas de bien caractériser $p(x_k|x_{k-1}^{(s)})$, ce qui conduit à un ensemble $\{x_{k-1}^{(i)}, \lambda_k^{(i)}\}$ qui n'est pas représentatif de la densité $p(x_{k-1}|z_{1:k})$. Par conséquent, le rééchantillonnage risque d'éliminer certaines particules qui une fois propagées selon la dynamique auraient été très vraisemblables, et à l'inverse de multiplier d'autres particules qui après prédiction se retrouvent en queue de la vraisemblance.

L'algorithme RBSSHSSIR proposé dans [Torma et al., 2003] et décrit dans la Table 2.8 du §2.2.3–B inclut un rééchantillonnage auxiliaire. Celui-ci vise, pour chaque particule, à préserver la cohérence du point de vue de la dynamique entre sa particule prédécesseur et ses composantes échantillonnées sur la base de la mesure. L'algorithme RBSSHSSIR admet alors une structure de type « filtre à particules auxiliaires », et peut être démontré selon des arguments analogues à ceux sous-tendant l'AUXILIARY. Nous donnons ci-dessous les grandes lignes d'une telle preuve, en rappelant que (i) le vecteur d'état à l'instant k s'écrit $x_k = (u_k', v_k', h_k')'$ et évolue selon la dynamique $p(x_k|x_{k-1}) = p(u_k, v_k|x_{k-1})p(h_k|x_{k-1})$, avec $p(h_k|x_{k-1}) = \delta(h_k - f_2(x_{k-1}))$, (ii) seule la « partie innovation » $(u_k', v_k')'$ intervient dans le lien état-mesure, *i.e.* $p(z_k|x_k) = p(z_k|u_k, v_k)$, (iii) u_k et v_k sont échantillonnés séquentiellement, respectivement selon une fonction d'importance $q(u_k|x_{k-1}, z_k) = q(u_k|z_k)$ ne dépendant que de la mesure z_k, et selon $q(v_k|u_k, x_{k-1}, z_k) = p(v_k|u_k, x_{k-1})$.

Similairement à (2.76), $p(x_k, s|z_{1:k})$ peut être développé en

$$\begin{aligned} p(x_k, s|z_{1:k}) &= p(u_k, v_k, h_k, s|z_{1:k}) \\ &\propto p(z_k, u_k, v_k, h_k, s|z_{1:k-1}) \\ &\propto p(s|z_{1:k-1})p(u_k|s, z_{1:k-1})p(v_k|u_k, s, z_{1:k-1})p(h_k|u_k, v_k, s, z_{1:k-1})p(z_k|x_k, s, z_{1:k-1}) \end{aligned} \tag{2.84}$$

et satisfait ici

$$p(x_k, s|z_{1:k}) \propto w_{k-1}^{(s)} p(u_k|x_{k-1}^{(s)})p(v_k|u_k, x_{k-1}^{(s)})p(h_k|x_{k-1}^{(s)})p(z_k|u_k, v_k), \tag{2.85}$$

avec $p(h_k|x_{k-1}^{(s)}) = \delta(h_k - f_2(x_{k-1}^{(s)}))$. D'autre part, si on définit la fonction d'importance

$$\begin{aligned} q(x_k, s|z_{1:k}) &= q(u_k, v_k, h_k, s|z_{1:k}) \\ &= q(u_k|z_{1:k})q(s|u_k, z_{1:k})q(v_k|u_k, s, z_{1:k})q(h_k|u_k, s, v_k, z_{1:k}) \end{aligned} \tag{2.86}$$

de telle sorte que

$$q(u_k|z_{1:k}) = q(u_k|z_k) \tag{2.87}$$

$$q(s|u_k, z_{1:k}) = \frac{w_{k-1}^{(s)} p(u_k|x_{k-1}^{(s)})}{\sum_{r=1}^{N} w_{k-1}^{(r)} p(u_k|x_{k-1}^{(r)})} \tag{2.88}$$

$$q(v_k|u_k, s, z_{1:k}) = p(v_k|u_k, x_{k-1}^{(s)}) \tag{2.89}$$

$$q(h_k|u_k, s, v_k, z_{1:k}) = p(h_k|x_{k-1}^{(s)}), \tag{2.90}$$

alors le poids $w_k^{(i)}$ associé à chaque particule $(x_k^{(i)}, s^{(i)}) = (u_k^{(i)}, v_k^{(i)}, h_k^{(i)}, s^{(i)})$ s'écrit

$$\begin{aligned} w_k^{(i)} &\propto \frac{p(x_k^{(i)}, s^{(i)}|z_{1:k})}{q(x_k^{(i)}, s^{(i)}|z_{1:k})} \\ &\propto \frac{p(z_k|u_k^{(i)}, v_k^{(i)})}{q(u_k^{(i)}|z_k)} \sum_{r=1}^{N} w_{k-1}^r p(u_k^{(i)}|x_{k-1}^{(r)}). \end{aligned} \tag{2.91}$$

Bien sûr, comme indiqué au début du §2.3.1, l'ensemble de particules pondérées $\{(x_k^{(i)}, s^{(i)}), w_k^{(i)}\}$ représente $p(x_k, s|z_{1:k})$ et, par marginalisation, $p(x_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)})$.

Ces développements apportent un nouvel éclairage sur l'algorithme RBSSHSSIR présenté Table 2.8. L'échantillonnage de $(x_k^{(i)}, s^{(i)})$ est hiérarchisé, vu (2.86),(2.87),(2.88),(2.89),(2.90), en les échantillonnages successifs de $u_k^{(i)} \sim q(u_k|z_k)$, $s^{(i)} \sim \frac{w_{k-1}^{(s)} p(u_k^{(i)}|x_{k-1}^{(s)})}{\sum_{r=1}^{N} w_{k-1}^r p(u_k^{(i)}|x_{k-1}^{(r)})}$, $v_k^{(i)} \sim p(v_k|u_k^{(i)}, x_{k-1}^{s^{(i)}})$, $h_k^{(i)} \sim p(h_k|x_{k-1}^{s^{(i)}})$, qui correspondent respectivement aux items 6,7,8,9 de l'algorithme RBSSHSSIR. En outre, on reconnaît dans l'item 10 du même algorithme la formule de mise à jour des poids (2.91).

Notons enfin que le raisonnement peut être aisément étendu au cas où le vecteur d'état est limité à $x_k = (u_k', v_k')'$, *e.g.* lorsque les composantes de u_k et v_k suivent des dynamiques indépendantes du premier ordre Il permet de prouver que le RBSSHSSIR demeure valide, modulo le fait que $x_k^{(i)}$ soit seulement constitué de la superposition des sous-vecteurs $u_k^{(i)}$ et $v_k^{(i)}$ dans l'item 9 Table 2.8.

2.3.2 Stratégie Unscented

Le « filtre particulaire unscented » –UPF : *Unscented Particle Filter* [Merwe et al., 2000]– est une alternative aux mécanismes précédemment cités, consistant à approximer pour chaque particule $x_k^{(i)}$ la fonction d'importance optimale associée $q^*(x_k|x_{k-1}^{(i)}, z_k) = p(x_k|x_{k-1}^{(i)}, z_k)$ –cf. (2.30)– par une Gaussienne, de sorte que $x_k^{(i)} \sim q(x_k|x_{k-1}^{(i)}, z_k) = \mathcal{N}(x_k; m_{k|k}^{(i)}, P_{k|k}^{(i)})$. Pour chaque indice i, les moments $(m_{k|k}^{(i)}, P_{k|k}^{(i)})$ sont établis à partir des moments $(m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)})$ –relatifs à la Gaussienne selon laquelle est échantillonnée la « particule prédécesseur » $x_{k-1}^{(i)}$– et de l'observation z_k, au moyen d'une extension du filtre de Kalman appelée « filtre de Kalman unscented » –UKF : *Unscented Kalman Filter* [Julier et al., 1997].

Nous présentons ci-dessous la *transformée unscented*, qui constitue le fondement théorique de l'UKF. Nous donnons ensuite un bref aperçu de l'algorithme de l'UKF. Enfin, nous concluons par l'algorithme de l'UPF[7].

[7]Signalons qu'il existe un « filtre particulaire étendu » –EPF : *Extended Particle Filter*–, qui se décline de manière semblable à l'UPF excepté le fait que les moments $(m_{k|k}^{(i)}, P_{k|k}^{(i)})$ relatifs à la loi de proposition de $x_k^{(i)}$ sont obtenus à partir de $(m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)})$ et z_k par application d'un pas du filtre de Kalman étendu.

A Transformée unscented

La « transformée unscented » [Julier et al., 1997] est une méthode de calcul approché des moments de l'image d'une variable aléatoire $x \in \mathbb{R}^{n_x}$ par une transformation non linéaire $f(.)$. Elle consiste à faire subir cette transformation à $2n_x + 1$ "σ-points" $\mathcal{X}_0, \ldots, \mathcal{X}_{2n_x}$, judicieusement sélectionnés de manière déterministe selon la distribution de probabilité $p_x(x)$ de x, et à approximer la densité de probabilité de sortie en se basant uniquement sur les images $f(\mathcal{X}_0), \ldots, f(\mathcal{X}_{2n_x})$ de ces points, cf. Figure 2.3.

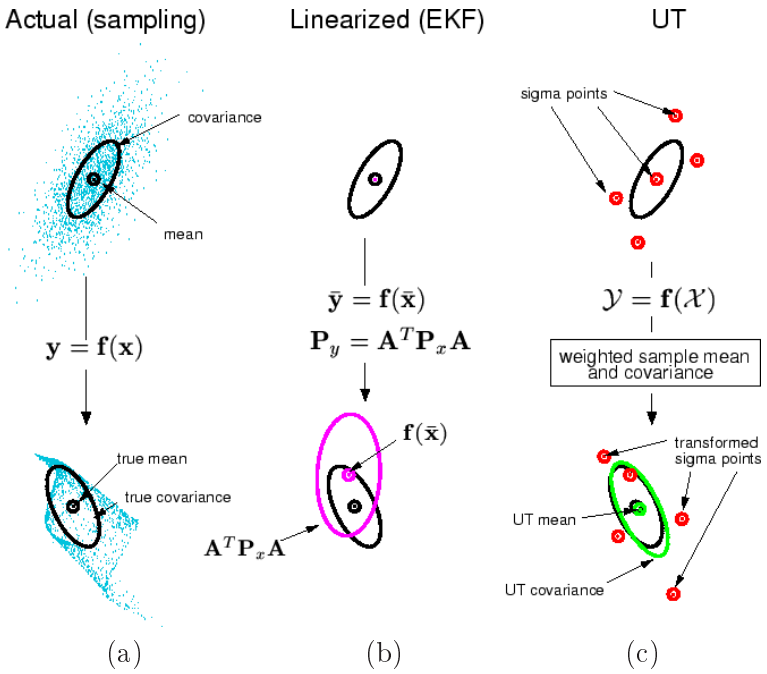


FIG. 2.3 – Exemple de propagation des moyennes et des covariances d'une distribution au travers d'une transformation non linéaire (extrait de [Merwe et al., 2000]). (a) Valeurs réelles des moments obtenues par échantillonnage. (b) Propagation par linéarisation au premier ordre (méthode utilisée dans l'EKF). (c) Propagation à l'aide des σ-points (transformée unscented, utilisée dans l'UKF).

Lorsque $\mathbb{E}_{p_x(.)}(x) = \bar{x}$ et $\mathbb{E}_{p_x(.)}(x - \bar{x})(x - \bar{x})' = P_x$, les "$\sigma$-points" peuvent être définis par

$$\begin{aligned} \mathcal{X}_0 &= \bar{x}\,, \\ \mathcal{X}_i &= \bar{x} + \left(\sqrt{(n_x + \lambda)P_x}\right)_i, \; i = 1, \ldots, n_x\,, \\ \mathcal{X}_i &= \bar{x} - \left(\sqrt{(n_x + \lambda)P_x}\right)_{i-n_x}, \; i = n_x + 1, \ldots, 2n_x, \end{aligned} \tag{2.92}$$

où $(\sqrt{M})_i$ désigne la $i^{\text{ème}}$ colonne de la matrice triangulaire inférieure de Cholesky N telle que $NN' = M$, et où $\lambda = \alpha^2(n_x + \kappa) - n_x$ est un paramètre de dimensionnement.

La constante α détermine l'étendue de la répartition des σ-points autour de $\bar{x}$, et satisfait généralement $\alpha \in [10^{-4}; 1]$. Le paramètre de dimensionnement secondaire κ est généralement compris entre 0 et $3 - n_x$, cf. [Julier, 2002] pour davantage de détails.

Soient

$$\mathcal{Y}_i = f(\mathcal{X}_i),\ i = 0, \ldots, 2n_x \tag{2.93}$$

les images des σ-points par $f(.)$. Les deux premiers moments de la variable aléatoire $y = f(x)$ sont alors approximés par

$$\overline{f(x)} = \mathbb{E}_{p_x(.)}(f(x)) \approx \bar{y} = \sum_{i=0}^{2n_x} W_i^{(m)} \mathcal{Y}_i \tag{2.94}$$

$$\mathbb{E}_{p_x(.)}(f(x) - \overline{f(x)})(f(x) - \overline{f(x)})' \approx P_y = \sum_{i=0}^{2n_x} W_i^{(c)} (\mathcal{Y}_i - \bar{y})(\mathcal{Y}_i - \bar{y})', \tag{2.95}$$

où les poids $W_0^{(m)}$, $W_0^{(c)}$, $W_i^{(m)}$, $W_i^{(c)}$ satisfont

$$\begin{aligned} W_0^{(m)} = \frac{\lambda}{(n_x + \lambda)},\ W_0^{(c)} = \frac{\lambda}{(n_x + \lambda)} + (1 - \alpha^2 + \beta), \\ W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n_x + \lambda)},\ i = 1, \ldots, 2n_x. \end{aligned} \tag{2.96}$$

Le scalaire β permet de prendre en compte une connaissance *a priori* concernant la distribution de x. Si x suit une loi Gaussienne, alors on fixe $\beta = 2$ cf. [Merwe et al., 2000].

Ce processus permet une approximation de la moyenne et de la covariance de la variable aléatoire $f(x)$ précise jusqu'au deuxième ordre de leurs développements de Taylor (jusqu'au 3^e ordre dans le cas Gaussien), la précision des moments d'ordres supérieurs ou égaux à 3 étant conditionnée par les choix de α et β.

B Filtre de Kalman unscented

Comme précédemment mentionné, le filtre de Kalman unscented est une alternative au filtre de Kalman étendu pour l'approximation des deux premiers moments de la densité *a posteriori* $p(x_k|z_{1:k})$ du vecteur d'état d'un système non linéaire à partir de la connaissance des deux moments associés à $p(x_{k-1}|z_{1:k-1})$ et de la mesure z_k. Ce filtre admet une structure prédiction / mise à jour classique, et procède donc en deux temps. La phase de prédiction consiste à approximer les moments de la densité de prédiction $p(x_k|z_{1:k-1})$ par application de la transformée unscented sur des σ-points échantillonnés selon $p(x_{k-1}|z_{1:k-1})$ ainsi que selon le bruit de dynamique. Les équations de mise à jour peuvent quant à elles être justifiées en rappelant que la distribution *a posteriori* $p(x_k|z_{1:k})$ est égale à la distribution du vecteur d'état prédit à l'instant k conditionnée sur le fait que le prédicteur de la mesure se réalise en l'observation z_k. Ainsi, le calcul de la moyenne et de la covariance *a posteriori* nécessite, outre la donnée de z_k et des statistiques du prédicteur $x_k|z_{1:k-1}$, le calcul de la moyenne du prédicteur de la sortie $z_k|z_{1:k-1}$, de sa covariance, ainsi que de l'intercovariance entre $x_k|z_{1:k-1}$ et $z_k|z_{1:k-1}$,

ces trois dernières quantités étant approximées au moyen de la transformée unscented[8]. Comme pour le filtre de Kalman, une version « racine carrée » de l'UKF – SRUKF : *Square-Root Unscented Kalman Filter* – a été développée [Merwe et al., 2001] de façon à s'affranchir du mauvais conditionnement numérique des équations « classiques ». Celle-ci présente en outre une complexité significativement moindre.

C Filtre particulaire unscented

L'association des moments $(m^{(i)}_{k-1|k-1}, P^{(i)}_{k-1|k-1})$ à chaque particule $x_{k-1}(i)$, leur propagation entre deux instants selon un pas d'UKF tenant compte de la mesure z_k, et l'échantillonnage de $x^{(i)}_k$ selon $\mathcal{N}(x_k; m^{(i)}_{k|k}, P^{(i)}_{k|k})$ interviennent donc dans l'algorithme du filtre particulaire unscented détaillé Table 2.10. On reconnaît une structure comparable à l'algorithme générique SIR présenté Table 2.3 page 30.

2.3.3 Stratégie mixte

Le rééchantillonnage intermédiaire constitue l'étape fondamentale du filtre à particules auxiliaires décrit au §2.3.1, cf. les items 5–10 de l'algorithme AUXILIARY, Table 2.9 page 49. Il a pour but de sélectionner et multiplier, préalablement à leur propagation selon la dynamique du système, les particules dont il est pressenti que le « futur » couvrira des zones de l'espace d'état fortement vraisemblables vis à vis de l'observation. Cependant, il a été vu que la définition de l'approximation $\hat{p}(z_k|x_{k-1})$ de $p(z_k|x_{k-1})$, sur laquelle repose ce rééchantillonnage intermédiaire, est délicate. Ainsi, approximer $p(z_k|x^{(i)}_{k-1})$ par une fonction $p(z_k|\mu^{(i)}_k)$, où $\mu^{(i)}_k$ désigne une caractéristique – *e.g.* mode, espérance, échantillon – de $p(x_k|x^{(i)}_{k-1})$, conduit à un comportement médiocre du filtre dès lors que la dynamique est très diffuse ou si la vraisemblance $p(z_k|x_k)$ varie significativement lorsque x_k décrit les zones fortement probables du point de vue de $p(x_k|x^{(i)}_{k-1})$.

L'utilisation de la transformée unscented permet d'aboutir à une meilleure approximation de $p(z_k|x_{k-1})$. Il suffit pour cela de remarquer, d'après (2.32), que pour chaque particule $x^{(i)}_{k-1}$, $p(z_k|x^{(i)}_{k-1})$ est en fait l'espérance de l'image par la fonction $x \mapsto p(z_k|x)$, connue a une constante près d'une variable aléatoire x se distribuant selon $p(x|x^{(i)}_{k-1})$. On peut en outre décider de définir les σ-points relatifs à $p(x|x^{(i)}_{k-1})$ en prenant également en compte le fait qu'une distribution Gaussienne est associée à chaque particule

[8] Une écriture rigoureuse nécessite la différenciation entre une variable aléatoire X et sa réalisation x. Du fait que $p(X_k|Z_{1:k} = z_{1:k}) = p\big((X_k|Z_{1:k-1} = z_{1:k-1})|(Z_k|Z_{1:k-1} = z_{1:k-1}) = z_k\big)$, il vient les formules classiques $\mathbb{E}(X_k|Z_{1:k} = z_{1:k}) = \mathbb{E}(X_k|Z_{1:k-1} = z_{1:k-1}) + K_k\big(z_k - \mathbb{E}(Z_k|Z_{1:k-1} = z_{1:k-1})\big)$ et $\mathrm{Cov}(X_k|Z_{1:k} = z_{1:k}) = \mathrm{Cov}(X_k|Z_{1:k-1} = z_{1:k-1}) + K_k P_{Z_k|Z_{1:k-1},Z_k|Z_{1:k-1}} K_k^{'}$, avec $K_k = P_{X_k|Z_{1:k-1},Z_k|Z_{1:k-1}} P^{-1}_{Z_k|Z_{1:k-1},Z_k|Z_{1:k-1}}$, les notations $P_{A,B}$ et $P_{B,B}$ désignant respectivement l'intercovariance des variables aléatoires A et B, ainsi que la covariance de B. Le calcul approché de $\mathbb{E}(X_k|Z_{1:k} = z_{1:k})$ et $\mathrm{Cov}(X_k|Z_{1:k} = z_{1:k})$ à partir des moments – approximatifs – $\mathbb{E}(X_k|Z_{1:k-1} = z_{1:k-1})$ et $\mathrm{Cov}(X_k|Z_{1:k-1} = z_{1:k-1})$ nécessite donc la connaissance de $\mathbb{E}(Z_k|Z_{1:k-1} = z_{1:k-1})$, $P_{X_k|Z_{1:k-1},Z_k|Z_{1:k-1}}$, et $P_{Z_k|Z_{1:k-1},Z_k|Z_{1:k-1}}$. Ce sont précisément ces dernières quantités qui sont approximées au moyen de la transformée unscented.

$\left[\{x_k^{(i)}, w_k^{(i)}\}\right]_{i=1}^{N} = \text{UPF}([\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^{N}, z_k)$

SI $k = 0$ **(INITIALISATION) ALORS**
  POUR $i = 1, \ldots, N$, **FAIRE**
    Échantillonner $x_0^{(i)} \sim p(x_0)$ et poser $w_0^{(i)} = \frac{1}{N}$
    Initialiser les moyennes $m_{0|0}^{(i)}$ et covariances $P_{0|0}^{(i)}$ associées à $x_0^{(i)}$, $i = 1, \ldots, N$
  FIN POUR
FIN SI
SI $k \geq 1$ **ALORS**
  POUR $i = 1, \ldots, N$, **FAIRE**
    Mettre à jour les moments de la fonction d'importance associée à la particule $x_k^{(i)}$

$$[m_{k|k}^{(i)}, P_{k|k}^{(i)}] = \text{UKF}\left(m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}, z_k\right)$$

    Échantillonner $x_k^{(i)} \sim q(x_k | x_{0:k-1}^{(i)}, z_{1:k}) = \mathcal{N}(x_k; m_{k|k}^{(i)}; P_{k|k}^{(i)})$
    Mettre à jour le poids $w_k^{(i)}$ selon l'équation

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{0:k-1}^{(i)}, z_{1:k})}$$

    Normaliser les poids d'importance

$$w_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}}$$

    de telle sorte que $\sum_{i=1}^{N} w_k^{(i)} = 1$
  FIN POUR
FIN SI

TAB. 2.10 – Algorithme de filtrage particulaire unscented

$x_{k-1}^{(i)}$, dans la veine de l'UPF.

Andrieu *et al.* proposent une telle stratégie dans [Andrieu et al., 2001]. L'algorithme AUXILIARY_UNSCENTED – *Auxiliary Unscented Particle Filter* – ainsi obtenu est résumé Table 2.11. Il permet donc à la fois la définition d'une fonction d'importance $q(x_k|x_{0:k-1}, z_k)$ et une meilleure approximation de l'ensemble de coefficients $\{\hat{p}(z_k|x_{k-1}^{(i)})\}$ intervenant dans le rééchantillonnage intermédiaire. Notons toutefois que malgré son attrait et sa proximité du cas optimal, cette stratégie est plus difficile à mettre en œuvre et de complexité algorithmique plus élevée.

2.4 Synthèse

Nous avons présenté dans ce chapitre des méthodes de filtrage particulaire. Après avoir rappelé quelques généralités sur le filtrage particulaire mettant en œuvre les méthodes de Monte Carlo nous introduisons l'algorithme dit d'échantillonnage pondéré séquentiel. Il permet de construire récursivement un nuage de particules pondérées approchant la loi de filtrage. Toutefois, sa nature récursive conduit à une dégénérescence

$\left[\{x_k^{(i)}, w_k^{(i)}\}\right]_{i=1}^{N} = \text{AUXILIARY\_UNSCENTED}(\left[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}\right]_{i=1}^{N}, z_k)$

1: **SI** $k = 0$ **(INITIALISATION) ALORS**
2: **POUR** $i = 1, \ldots, N$, **FAIRE**
3: Échantillonner $x_0^{(i)} \sim p(x_0)$ et poser $w_0^{(i)} = \frac{1}{N}$
4: Initialiser les moyennes $m_{0|0}^{(i)}$ et covariances $P_{0|0}^{(i)}$ associées à $x_0^{(i)}$, $i = 1, \ldots, N$
5: **FIN POUR**
6: **FIN SI**
7: **SI** $k \geq 1$ **ALORS**
8: **POUR** $i = 1, \ldots, N$, **FAIRE**
9: Calculer les poids auxiliaires $\lambda_k^{(i)}$ selon

$$\lambda_k^{(i)} = q(i|z_{1:k}) \propto w_{k-1}^{(i)} \hat{p}(z_k|x_{k-1}^{(i)})$$

où $\hat{p}(z_k|x_{k-1}^{(i)})$ est une approximation de $p(z_k|x_{k-1}^{(i)})$ obtenue à partir de la transformée unscented
10: **FIN POUR**
11: Rééchantillonner l'ensemble de particules et de statistiques associées $\{x_{k-1}^{(i)}, m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}\}_{i=1}^{N}$ proportionnellement aux poids $\lambda_k^{(i)}$; renommer $\{(x_{k-1}^{(i)}, m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}); \frac{1}{N}\}$ l'ensemble équipondéré ainsi obtenu
12: **POUR** $i = 1, \ldots, N$, **FAIRE**
13: Mettre à jour les moments de la fonction d'importance associée à la particule $x_k^{(i)}$

$$[m_{k|k}^{(i)}, P_{k|k}^{(i)}] = UKF\left(m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}, z_k\right)$$

14: Échantillonner $x_k^{(i)} \sim q(x_k|x_{0:k-1}^{(i)}, z_{1:k}) = \mathcal{N}(x_k; m_{k|k}^{(i)}, P_{k|k}^{(i)})$
15: Mettre à jour le poids $w_k^{(i)}$ selon l'équation

$$w_k^{(i)} \propto \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{\hat{p}(z_k|x_{k-1}^{(i)})q(x_k^{(i)}|x_{0:k-1}^{(i)}, z_{1:k})}$$

16: **FIN POUR**
17: **FIN SI**

TAB. 2.11 – Algorithme de filtrage unscented avec poids auxiliaires (AUXILIARY_UNSCENTED)

du nuage induit par l'augmentation dans le temps de la variance inconditionnelle de ses poids. Nous détaillons ensuite la fonction d'importance dite optimale qui permet de positionner les particules en tenant compte de la dynamique du système et de l'observation à l'instant courant. On montre que pour une telle fonction, le phénomène de dégénérescence est compensé mais en pratique, cette fonction n'est généralement pas utilisable. D'autre stratégies sont alors envisagées. L'introduction d'un rééchantillonnage dans l'algorithme, permet de limiter la dégénérescence des poids d'importance en éliminant les particules de poids négligeables et en multipliant les particules les plus significatives. Le filtre résultant constitue l'algorithme générique de filtrage à partir duquel toute les stratégies de filtrage particulaire dérivent. D'autres méthodes telle que l'échantillonnage partitionné permettant de réduire la variance des poids d'importance sont aussi présen-

tées. Enfin différents choix de fonction d'importance sont discutés. Parmi les fonctions d'importance dites "simples", l'utilisation par exemple d'une fonction d'importance exclusivement basée sur la dynamique du système conduit à l'algorithme bien connu de CONDENSATION qui est abondamment utilisé en suivi visuel. Enfin d'autres fonctions d'importances plus complexes permettant de prendre en compte l'observation dans le positionnement des particules et ainsi de s'approcher du cas optimal sont présentées. La stratégie "Auxiliary" par exemple, utilise des poids auxiliaires pour selectionner les particules les plus vraisemblables avant leur propagation tandis que le filtre particulaire unscented utilise la transformée unscented à travers un filtre de Kalman unscented pour prendre en compte à la fois la dynamique du système et l'observation courante dans la propagation des particules.

Malgré le nombre de stratégies proposées, ce chapitre n'est pas complètement exhaustif, d'autres stratégies de suivi restent envisageables. Nous n'avons pas non plus considéré les algorithmes de suivi multi-cible. Dans ce contexte, Kotecha et Djuric dans [Kotecha et al., 2003] d'une part et van der Merwe dans [van der Merwe et al., 2003] d'autre part representent la densité *a posteriori* par un mélange de Gaussiennes afin de maintenir la multimodalité.

Chapitre 3

Attributs visuels pour le filtrage particulaire

3.1 Introduction

Dans le chapitre précédent, nous avons présenté les méthodes de filtrage particulaire mises en œuvre dans notre travail. Les mesures jouent un rôle essentiel dans le fonctionnement du filtre, d'une part dans la définition d'une fonction de vraisemblance des particules, et d'autre part dans la définition d'une fonction d'importance qui détermine la stratégie d'exploration de l'espace d'état. Elles sont extraites des images acquises dans le flot vidéo. L'information contenue dans une image étant très riche, des attributs de natures diverses, typiquement de forme, couleur et mouvement, peuvent être considérés. Comme Pérez *et al.* dans [Pérez et al., 2004], nous les classifierons en attributs persistants ou intermittents selon le contexte applicatif. Les attributs persistants permettent d'obtenir une mesure systématique mais souvent peu discriminante, par exemple un attribut de forme dans un environnement très encombré. Les attributs intermittents dans le flot vidéo sont par nature discriminants ; ils sont souvent issus de modules de détection, éventuellement combinés avec des attributs persistants pour caractériser, à chaque instant, l'état de la cible suivie.

Les sections 3.2 à 3.4 décrivent successivement les attributs proposés de mouvement, couleur et forme, ainsi que leurs fonctions d'importance et de mesure associées. Comme énoncé dans le chapitre précédent, nous rappelons que certains algorithmes de filtrage particulaire utilisent une fonction d'importance (section 2.2) pour guider l'exploration de l'espace d'état. La fonction de mesure joue, quant à elle, un rôle essentiel dans le filtre car les vraisemblances qu'elle définit doivent permettre à l'ensemble de particules pondérées de représenter correctement la distribution *a posteriori*. La fonction de mesure doit être discriminante et calculable à chaque instant image. De fait, elle intègre au moins un attribut visuel persistant. La section 3.5 propose une évaluation de telles fonctions en terme de pouvoir discriminant, précision et temps de calcul lorsque les attributs sont pris séparément, combinés ou fusionnés.

Dans les deux derniers chapitres, ces fonctions et surtout leurs associations envisa-

gées dans diverses stratégies de filtrage particulaire sont discutées et évaluées dans un contexte de suivi de personnes ou de reconnaissance de gestes.

3.2 Attribut mouvement

3.2.1 Généralités

L'analyse de séquences d'images permet la définition d'attributs visuels rendant compte du mouvement inter-image des régions mobiles dans la scène [Konrad, 2000]. Nous supposerons ici la caméra immobile et par conséquent le robot à l'arrêt et les actionneurs de la platine figés. Sous cette hypothèse, il est alors trivial d'obtenir une information sur le mouvement de la scène. Certes, il existe des techniques, par exemple [Batista et al., 1998], qui permettent une segmentation de la scène par le mouvement à partir d'un capteur non statique mais leur mise en œuvre sur des plateformes robotiques reste relativement coûteuse en temps de calcul. Pour une caméra non statique, d'autres attributs autres que le mouvement seront considérés.

Le mouvement est par nature souvent intermittent, ce qui impose, en général, de ne pas l'utiliser seul dans une fonction de mesure sauf si le mouvement image est supposé persistant (section 3.2.3). Il est aussi intéressant pour la définition d'une fonction d'importance (section 3.2.2). Pour caractériser ce mouvement image, nous utilisons classiquement le flot optique et la différence absolue entre images successives, qui sont rappelés ci-après.

Le *flot optique*, ou *champ dense de vitesse apparente 2D*, est régi par l'équation de contrainte de mouvement :

$$\frac{\partial I}{\partial t}(x,y,t) + \frac{\partial I}{\partial x}(x,y,t).u + \frac{\partial I}{\partial y}(x,y,t).v = 0 \tag{3.1}$$

où $I(x,y,t)$ représente l'intensité à l'instant t du pixel de coordonnées (x,y) et $\overrightarrow{U} = (u,v)'$ désigne le vecteur vitesse apparent. Ce vecteur ne peut pas être estimé par la seule équation (3.1) d'où l'ajout de contraintes supplémentaires, par exemple dans [Horn et al., 81].

Bien que l'étude du flot optique permette d'extraire une information sur la direction du déplacement des objets, nous ne nous intéressons ici qu'au module du déplacement pour construire la mesure z^{M_f} (avec M pour Mouvement et f pour flot optique). On obtient alors une image binaire labélisant les pixels en mouvement.

Deux alternatives sont alors possibles pour prendre en compte cette information :

1. effectuer un étiquetage spatial des pixels pour segmenter les régions mobiles, notées R_x ; ces *blobs* sont adaptés pour définir une fonction d'importance ;
2. combiner dans la fonction de mesure le flot optique calculé en chaque pixel avec d'autres attributs.

La figure 3.1 montre deux images successives d'une séquence et l'amplitude du flot optique calculé. En marge du flot optique, on peut exploiter la différence absolue entre

(a) (b) (c)

FIG. 3.1 – Un exemple de flot optique calculé : (a) image à l'instant $k-1$, (b) image à l'instant k, (c) amplitude du flot optique

l'image courante et une image de référence, classiquement l'image précédente du flot vidéo. Si on appelle I_k l'image à l'instant k, alors

$$z^{M_d}(i,j) = |I_k(i,j) - I_{k-1}(i,j)|.$$

La figure 3.2-(a) montre l'image différence correspondant aux deux images successives figure 3.1-(a) et figure 3.1-(b)

Nous reprenons ici le détecteur de mouvement défini par Pérez *et al.* dans [Pérez et al., 2004] qui repose sur l'image différence z^{M_d}. L'activité en terme de mouvement d'une région rectangulaire R_x de position image et dimensions (échelle) données est caractérisée à partir de z^{M_d} par comparaison avec un histogramme $h_{i,x}$ sur N_{bi} cellules indexées par i et tel que

$$h_{i,x} = K \sum_{u \in R_x} \delta_i(b_u), i = 1, \ldots, N_{bi} \tag{3.2}$$

où $b_u \in \{1 \ldots N_{bi}\}$ est l'index de la cellule incluant l'intensité du pixel $u \in R_x$, δ_a désigne la fonction de Kronecker en a, et K est un coefficient de normalisation tel que $\sum_{i=1}^{N_{bi}} h_{i,x} = 1$.

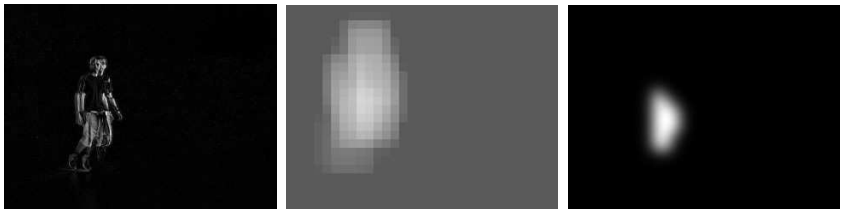
Chaque histogramme $h_{i,x}$ doit être comparé à un histogramme de référence afin de conclure si la région R_x est ou non en mouvement. Pérez *et al.* dans [Pérez et al., 2004] ont montré que le choix d'une distribution uniforme pour histogramme de référence $h_{i,ref}$ est un bon compromis pour caractériser une région mobile. Par conséquent :

$$h_{i,ref} = \frac{1}{N_{bi}},\ i = 1, \ldots, N_{bi}$$

Ce modèle correspond à une répartition du mouvement dans tous les niveaux et indique donc que la zone n'est pas statique. L'histogramme $h_{i,x}$ d'une région candidate peut alors être comparé à la référence pour déterminer si cette dernière est mobile. Comme dans [Comaniciu et al., 2003] et [Pérez et al., 2004], nous utilisons une distance de Bhattacharyya, soit :

$$D(h_x, h_{ref}) = \left(1 - \sum_{i=1}^{N_{bi}} \sqrt{h_{i,x}, h_{i,ref}}\right)^{\frac{1}{2}}. \tag{3.3}$$

La figure 3.2.(b) montre la valeur en chaque pixel de la distance D pour l'exemple de la figure 3.1.



(a) (b) (c)

FIG. 3.2 – Extraction des régions par différence d'images : (a) image différence z^{M_d}, (b) valeurs de la distance de Bhattacharyya dans chaque sous région de l'image, (c) fonction d'importance associée

3.2.2 Fonctions d'importance associées

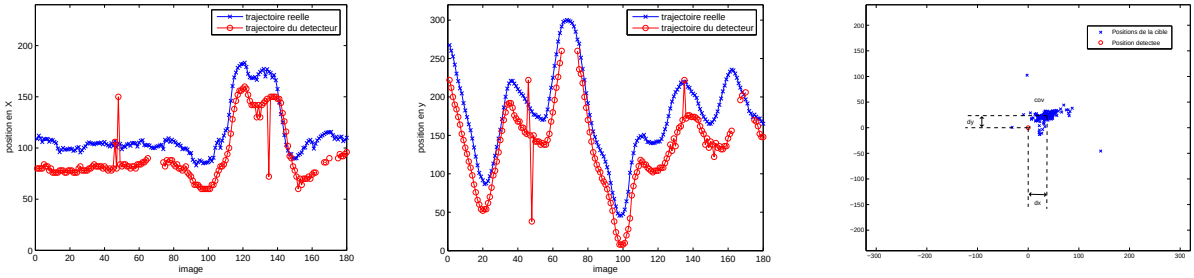
Le calcul de la distance D étant coûteux en temps de calcul, celle-ci est calculée sur des régions R_x de taille fixe et positionnées à intervalles réguliers (typiquement tous les 10 pixels). Soient x_{pos} ces coordonnées image. Un seuillage sur cette distance i.e. $D^2(h_{x_{pos}}, h_{ref}) > \tau$ permet ensuite d'obtenir les positions des B régions R_x de forte activité. Empiriquement, nous avons fixé $\tau = 0.75$.

A partir de ces B *blobs* détectés par différence d'images ou flot optique, on peut définir une fonction d'importance, notée $q(x_k|z_k^M)$, concernant les composantes position du vecteur d'état. Elle se caractérise par un mélange de gaussiennes de covariances identiques centrées, à un décalage près, sur les positions des B *blobs* détectés [Isard et al., 1998b]. En effet, on retrouve dans la plupart des detecteurs de blobs un biais systématique qui doit être pris en compte dans la définition de la fonction d'importance. Ce biais correspond généralement à un décalage constant de la position détectée à partir d'un attribut vis à vis du modèle de la cible. Pour le mouvement par exemple, la périphérie de la cible est mieux détectée que son centre ce qui peut générer des décalages. Soit x_{pos_i} la position du *blob* détecté i. La fonction $q(x_k|z_k^M)$ s'écrit

$$q(x_k|z_k^M) = \sum_{i=1}^{B} \delta_i \mathcal{N}(b_i, \Sigma_B) \tag{3.4}$$

avec $b_i = x_{pos_i} + \bar{x}_B$. Les paramètres $\bar{x}_B$ et Σ_B désignent respectivement la moyenne et la covariance du décalage entre la position détectée et la position réelle de la cible.

Ces deux paramètres sont appris hors-ligne par comparaison, sur des séquences test, des positions détectées et des « vraies » positions sélectionnées manuellement. En traçant dans un repère centré sur la position détectée à chaque instant k les vraies positions du centre de la région d'intérêt, on obtient la distribution $\mathcal{N}(\bar{x}_B, \Sigma_B)$ comme illustré sur la figure 3.3. La figure 3.2.(c) montre la fonction d'importance pour l'exemple de la figure 3.1.



(a) (b) (c)

FIG. 3.3 – Estimation du décalage moyen et de la covariance du détecteur en fonction des positions réelles : (a) trajectoire en X de la cible et de la détection associée, (b) trajectoire en Y de la cible et de la détection, (c) caractérisation de la gaussienne à partir du tracé des vraies positions centrées sur la détection

3.2.3 Fonctions de mesure associées

Rappelons que l'activité d'une région R_x peut se quantifier par la distance 3.3. Sous l'hypothèse – peu réaliste dans notre contexte – d'un mouvement inter-image persistant, nous pouvons définir une fonction de mesure $p(z^{M_d}|x_k)$ basée sur le mouvement, soit avec les conventions usuelles,

$$p(z^{M_d}|x_k) \propto \exp\left(-\frac{D^2(h_x, h_{ref})}{2.\sigma_M^2}\right). \tag{3.5}$$

Des fonctions de mesure multi-attributs incluant le mouvement, en particulier le flot optique, seront présentées en section 3.4.7.

3.3 Attribut couleur

3.3.1 Généralités

La couleur est un attribut persistant et souvent caractéristique de la cible observée. Lors du suivi, nous pouvons ainsi exploiter une simple classification des pixels couleur, typiquement des pixels peau, *via* une segmentation en régions caractéristiques. Cette segmentation est notamment appropriée à la détection des cibles à suivre. Une variante consiste en une signature colorimétrique de la cible *e.g.* en sa distribution locale de couleur. Cet attribut reste néanmoins inadapté pour une étape de détection car trop coûteux en temps de calcul. Ces différents attributs colorimétriques sont décrits dans les trois sections suivantes pour des stratégies de filtrage dédiées. Ils permettent l'élaboration de diverses fonctions d'importance et de mesure décrites respectivement dans les sections 3.3.4 et 3.3.5.

3.3.2 Classification des pixels par leurs couleurs

La classification des pixels par leurs couleurs, notée z^{C_p} (avec C pour couleur et p pour peau), n'est possible que si nous disposons d'une connaissance *a priori* de la couleur de la cible. L'objectif étant de réaliser un suivi de certains membres corporels humains, un apprentissage de la couleur peau est ici envisageable. Chaque pixel image est alors affecté d'une probabilité d'appartenance à la classe peau. La carte de probabilité résultante dépend du choix de la base de couleur et de la méthode de classification choisie.

Les bases colorimétriques constituées des trois primaires R, V, B (standards CIE, FCC ou EBU) sont peu adaptées car elles restent trop sensibles aux variations d'illumination couramment rencontrées dans notre contexte applicatif. Il est préférable d'utiliser une base de couleur qui sépare les composantes luminance et chrominance pour ne conserver que les composantes chromatiques afin de représenter la couleur intrinsèque des objets. On distingue trois familles de bases colorimétriques selon la transformation appliquée à la base R, V, B :

- par transformation linéaire ($YCrCb$, YIQ, YUV, $I_1I_2I_3, ...$) ;
- par transformation non-linéaire (HSV, HIS, $HLS, ...$) ;
- par transformation fortement non-linéaire ($CIE - Lab$, $CIE - Luv, ...$).

Pour la classification des pixels peau, des études comparatives entre ces espaces ont été menées, par exemple dans [Zarit et al., 1999] et [Phung et al., 2005]. Elles montrent clairement qu'aucune base n'est plus adaptée qu'une autre : les taux de classification des pixels peau dépendent du contexte et de la méthode de classification. Nous avons finalement opté pour la base (I_1, I_2, I_3) en raison de ses performances connues en terme de séparabilité des classes [Ohta et al., 1980] et pour son faible coût calculatoire comparativement aux transformations non linéaires. Cette base est définie par les équations suivantes :

$$I_1 = \frac{R+G+B}{3}$$

$$I_2 = \frac{R-B}{2}$$

$$I_3 = \frac{2G-R-B}{4}$$

Indépendamment de la base choisie, un modèle de représentation du sous-espace colorimétrique sélectionné ainsi qu'une règle de décision doivent être adoptés. Des modèles paramétriques tels que Gaussienne, mélanges de Gaussiennes [Terrillon et al., 2000] ou ellipses [Lee et al., 2002] conduisent à une représentation compacte du sous-espace colorimétrique considéré. De plus, ils permettent de généraliser et d'interpoler la distribution de couleur, palliant ainsi un apprentissage incomplet ou non représentatif de la vraie distribution. En contrepartie, ces modèles sont des représentations approximatives de la distribution réelle apprise [Jones et al., 1999].

Comme Schwerdt *et al.* dans [Schwerdt et al., 2000], notre représentation repose sur des histogrammes construits par apprentissage, permettant ainsi de n'utiliser aucun modèle explicite de la couleur, ainsi que sur une règle de décision de type Bayésienne.

Vezhnevets *et al.* ont montré dans [Vezhnevets et al., 2003] que le classifieur Bayésien offre le meilleur compromis en terme de taux de vrais et faux positifs obtenus. Le principe de la méthode est rappelé ci-après.

Si on note $\mathbf{C}$ la couleur d'un pixel de l'image, la règle de Bayes permet d'écrire pour chaque pixel

$$p(peau|\mathbf{C}) = p(\mathbf{C}|peau)\frac{p(peau)}{p(\mathbf{C})}. \tag{3.6}$$

Cette distribution peut se caractériser par deux histogrammes h_{total} et h_{peau}. En effet,

$$p(peau) = \frac{N_{peau}}{N_{total}},\ p(\mathbf{C}) = \frac{1}{N_{total}}h_{total}(\mathbf{C}),\ p(\mathbf{C}|peau) = \frac{1}{N_{peau}}h_{peau}(\mathbf{C}),$$

où N_{total} désigne le nombre total de pixels d'une base d'images représentatives des scènes observées et N_{peau} est le nombre de pixels de couleur peau dans cette même séquence. En remplaçant dans l'équation (3.6) on obtient après simplification

$$p(peau|\mathbf{C}) = \frac{h_{peau}(\mathbf{C})}{h_{total}(\mathbf{C})}.$$

Le calcul de ce rapport d'histogrammes pour chaque pixel d'une image permet ainsi d'obtenir une carte de probabilité. Les pixels de fortes (resp. faibles) probabilités ont des niveaux proches du blanc (resp. du noir) sur la figure 3.4.(b).

3.3.3 Segmentation en régions peau

La segmentation en régions peau, notée z^{C_r}, a pour but de déterminer les régions d'une image cohérentes à la fois spatialement et du point de vue de leur contenu. A partir de la carte de probabilité z^{C_p}, une pré-segmentation de l'image est effectuée en sélectionnant les pixels de z^{C_p} pour lesquels $p(peau|C)$ est supérieur à un seuil préalablement fixé, *e.g.* $p(peau|C) > 0.5$. La figure 3.4.(c) montre le seuillage de la carte de probabilité de la figure 3.4.(b).

(a) (b) (c)

FIG. 3.4 – Un exemple de classification de pixels : (a) image originale, (b) carte de probabilité, (c) carte de probabilité seuillée

Un algorithme d'étiquetage spatial, *e.g.* local séquentiel ou récursif, permet d'isoler les régions dans l'image pré-segmentée par regroupement des pixels seuillés connexes.

Cependant, certaines régions liées à l'arrière-plan peuvent être ainsi segmentées. La figure 3.5 montre deux exemples de scènes encombrées incluant des régions colorimétriquement proches de la peau. Le deuxième exemple montre un cas extrême où il est impossible d'isoler spatialement la main du fond par un quelconque algorithme d'étiquetage spatial.

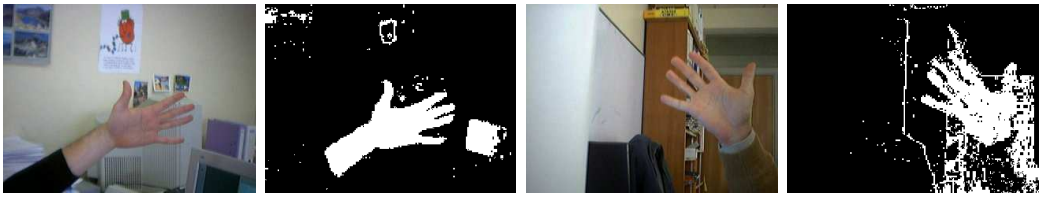


FIG. 3.5 – Deux images originales et les pré-segmentations associées

Durant cette thèse, des travaux préliminaires ont donc porté sur la segmentation région. Dans [Brèthes et al., 2004b], nous avons proposé une méthode de segmentation couleur qui combine la phase de pré-segmentation précédente à une phase de segmentation en régions par applications successives d'un algorithme de calcul de lignes de partage des eaux (LPE) [Beucher et al., 1993]. Cet algorithme permet le regroupement les pixels pré-segmentés connexes en régions cohérentes en termes de chrominance et luminance. Les régions peau sont ainsi mieux caractérisées tandis que les régions parasites sont filtrées grâce à des heuristiques relatives à leurs tailles, formes, etc.

Segmentation sur la chrominance

Un histogramme bidimensionnel sur les composantes (I_2, I_3) est tout d'abord généré à partir des pixels sélectionnés lors de la pré-segmentation. Considérant cet histogramme comme une image multi-niveaux de gris (figure 3.6(a)), nous appliquons un algorithme de LPE de façon à isoler les différents modes de couleurs présents dans l'image pré-segmentée. Afin de filtrer certains maxima locaux, cet histogramme est préalablement dilaté au moyen d'un élément structurant dont la taille correspond à la moyenne des écart-types calculés sur les composantes I_2 et I_3. Du fait que ceux-ci sont de l'ordre de $\sigma_{I_2} = 5.3$ et $\sigma_{I_3} = 2.5$ dans nos expérimentations, nous choisissons un élément structurant 4×4 (figure 3.6(b)). Nous appliquons alors notre algorithme de LPE. Celui-

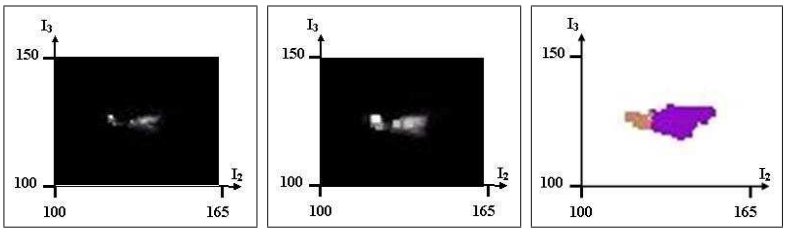


FIG. 3.6 – Histogrammes de la figure 3.5.(a) : (a) histogramme de chrominance, (b) histogramme dilaté, (c) histogramme clusterisé (deux regions sont segmentées)

ci requiert la sélection préalable de marqueurs. Leur nombre doit être judicieusement choisi car il correspond au nombre de régions finales à obtenir. Comme dans Albiol *et al.* [Albiol et al., 2001], nous calculons pour chaque mode de l'histogramme un *contraste normalisé* qui permet de filtrer les pics parasites non filtrés par la dilatation et ainsi éviter une sur-segmentation. Le contraste normalisé est défini pour chaque pic par

$$contraste\_normalise = \frac{Contraste}{Hauteur}$$

où $Hauteur$ désigne la hauteur du pic considéré et $Contraste$ est égal à la différence de hauteur entre ce pic et la vallée qui le sépare du pic voisin le plus élevé. La sélection des maxima est réalisée par simple seuillage de la fonction $contraste\_normalise$, typiquement un mode admettant un contraste normalisé supérieur à 10% est conservé et constitue un marqueur pour l'algorithme de LPE. La figure 3.7 montre les Hauteurs et Contrastes pour différents pics d'un histogramme. L'axe parallèle à l'axe des niveaux de gris identifie les marqueurs retenus.

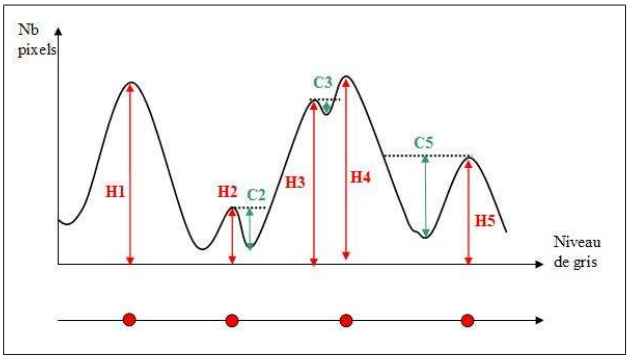


FIG. 3.7 – Hauteurs et contrastes pour les pics d'un histogramme

L'exécution de l'algorithme de LPE à partir des marqueurs ainsi définis permet de « clusteriser » l'histogramme de chrominance comme illustré sur la figure 3.6.(c). Les classes résultantes constituent les régions de z_r^C. La figure 3.8.(b) montre les régions ainsi obtenues sur les images vues précédemment. Dans le deuxième exemple, la couleur chair de l'armoire ne permet pas de segmenter proprement la main.

Segmentation par la luminance

La méthode reste la même que pour la chrominance mais nous l'appliquons sur les histogrammes d'intensité I_1 de chacune des régions précédemment caractérisées. La recherche des marqueurs s'effectue comme précédemment. Ainsi, l'image segmentée par la chrominance est ici éventuellement sur-segmentée à l'aide de la luminance.

Comme illustré sur la figure 3.8, cette nouvelle segmentation permet d'isoler la main dans le deuxième exemple mais reste sans effet sur le premier exemple qui présentait déjà une bonne segmentation.

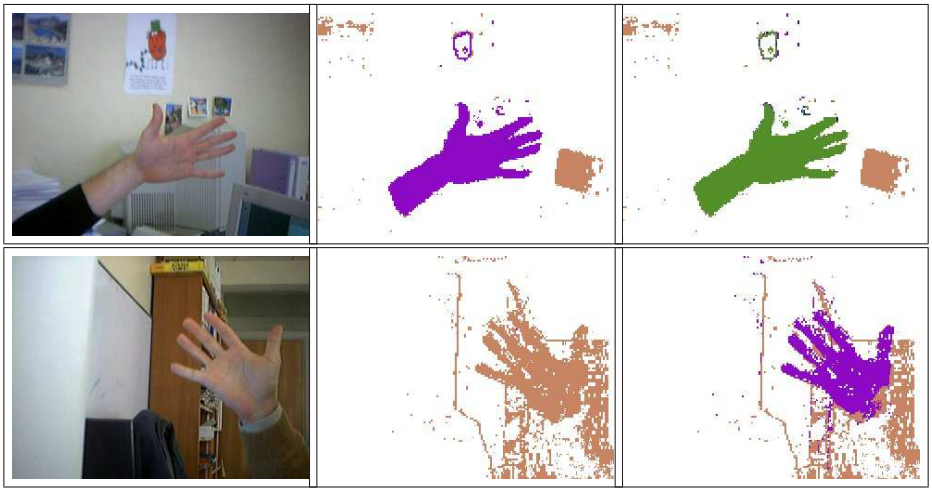


FIG. 3.8 – Deux exemples de segmentation : (a) images originales, (b) segmentations par chrominance, (c) segmentations par chrominance puis luminance. Les différentes couleurs correspondent aux regions segmentées

Élimination des petites régions

La dernière étape est de filtrer les régions inférieures à une taille donnée. Ce filtrage permet notamment d'éliminer des petites régions parasites connexes à la régions d'intérêt et isolées par les algorithmes successifs de LPE. La figure 3.9 montre quelques exemples de segmentation finale sur des scènes très encombrées.

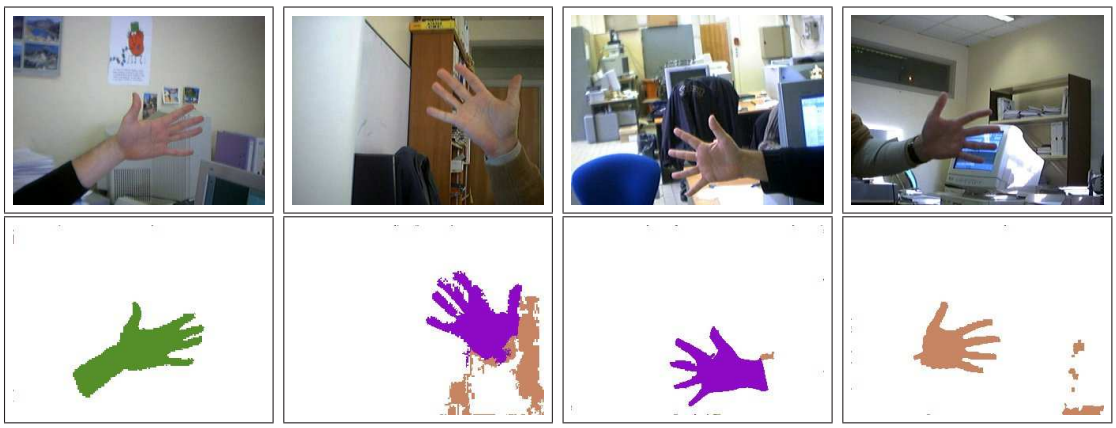
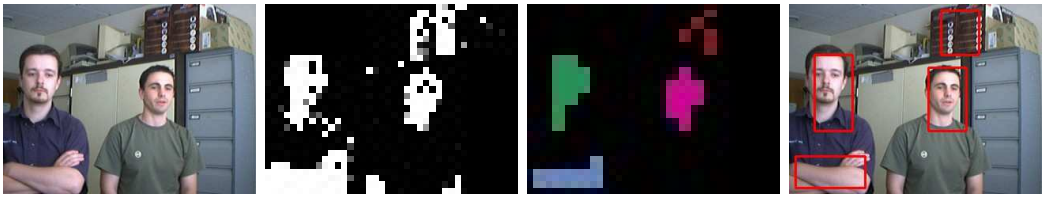


FIG. 3.9 – Quelques exemples de segmentation de régions peau pour des environnements encombrés

L'algorithme de segmentation couleur est évalué indirectement dans §3.5.

3.3.4 Fonctions d'importance associées

L'image traitée est tout d'abord sous-échantillonnée avec un facteur 10 afin d'obtenir une image de résolution plus faible. Une carte de probabilité peau z^{C_p} est ensuite calculée sur cette image et un filtre moyenneur de taille 2×2 est appliqué pour filtrer le bruit introduit par le sous-échantillonnage. Enfin, les pixels connexes sont regroupés par un algorithme d'étiquetage spatial avec un seuil hystérésis. Les régions ainsi segmentées sont définies par leurs rectangles englobants comme illustré sur la figure 3.10. Dans cet exemple résultant de la carte de probabilité z^{C_p}, on note la présence de fausses détections dans l'arrière-plan. Un traitement simple basé sur la forme des régions détectées, *e.g.* un test du rapport sur leurs dimensions, permet d'éliminer tout ou partie de ces fausses détections. Par analogie avec la fonction d'importance $q(x_k|z_k^M)$ sur le mouvement, nous



(a) (b) (c) (d)

FIG. 3.10 – Détection de *blobs* peau : (a) image originale, (b) carte de probabilité avant filtrage et seuillage, (c) régions obtenues après seuillage et filtrage, (d) résultat de la détection

définissons la fonction d'importance $q(x_k|z_k^C)$ basée sur la couleur comme un mélange de Gaussiennes correspondant aux B régions peau extraites et de positions x_{pos_i}. Avec les notations introduites en section 3.2.2, $q(x_k|z_k^C)$ s'écrit

$$q(x_k|z_k^C) = \sum_{i=1}^{B} \delta_i \mathcal{N}(b_i, \Sigma_B) \tag{3.7}$$

avec $b_i = x_{pos_i} + \bar{x}_B$. Les paramètres $\bar{x}_B$ et Σ_B, estimés par apprentissage, sont respectivement la moyenne et la covariance du décalage entre la position détectée et la position réelle de la cible.

La figure 3.11 montre sur un exemple la fonction d'importance $q(x_k|z_k^C)$ calculée en tout point image.

3.3.5 Fonctions de mesure associées

Nous reprenons et étendons la démarche introduite en section 3.2.3 en considérant ici des distributions locales sur les trois composantes de l'espace couleur. L'apparence de la cible dans l'image est représentée ici par trois histogrammes normalisés calculés sur la région d'intérêt R_x associée à cette cible. Pour chacun des trois plans, l'histogramme d'une région R_x est donné par la relation (3.2) que l'on généralise à trois canaux c :

$$h_{i,x}^c = c_K \sum_{u \in R_x} \delta_i(b_u^c), i = 1, \ldots, N_{bi},$$

(a) (b)

FIG. 3.11 – Calcul de fonction d'importance sur la couleur : (a) image originale, (b) fonction d'importance sur la couleur associée

où $b_u^c \in \{1 \dots N_{bi}\}$ indexe la cellule de l'histogramme correspondant au niveau du pixel u pour le canal c, et c_K est un terme de normalisation tel que $\sum_{i=1}^{N_{bi}} h_{i,x}^c = 1$. Les distributions de référence h_{ref}^c relatives à chaque plan sont calculées à partir d'une connaissance *a priori* sur la couleur de la cible suivie ou s'appuie sur la détection préalable de la cible à partir d'attributs autres, par exemple la détection de régions mobiles vue au § 3.2.2. La fonction de mesure (3.5) généralisée aux trois plans R, V, B devient

$$p(z^{C_{rvb}}|x) \propto \exp\left(-\sum_{c\in\{R,V,B\}} D^2(h_x^c, h_{ref}^c)/2\sigma_C^2\right). \tag{3.8}$$

La figure 3.12 montre le calcul de cette vraisemblance en tout point image avec un modèle de couleur (référence) correspondant au visage de la personne de droite. La région lui correspondant est donc logiquement privilégiée. La présence plausible de plusieurs individus nous amènera à considérer d'autres heuristiques dans le processus de suivi. Une

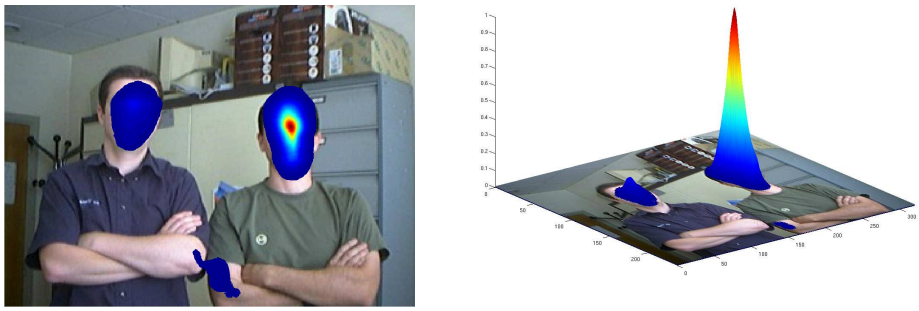


FIG. 3.12 – Exemple de vraisemblance en tout point image pour une mesure de distribution colorimétrique (une seule région d'intérêt)

extension logique est de considérer plusieurs régions d'intérêt distinctes spatialement et colorimétriquement [Nummiaro et al., 2002, Pérez et al., 2002]. Dans l'exemple 3.12, l'ajout d'une seconde distribution locale de couleur liée aux vêtements permettrait la

distinction des deux sujets. Plus globalement, la gestion de plusieurs sous-régions ou *patches* limite les dérives observées dans le temps, notamment lorsque la distribution de référence nécessite une mise à jour.

En généralisant à N_R sous régions et considérant $B_x = \bigcup_{p=1}^{N_R} B_{p,x}$ la région d'intérêt constituée de ces sous régions, le modèle de mesure (3.8) devient

$$p(z^{C_{rvb}}|x) \propto \exp\left(-\sum_{c\in\{R,V,B\}}\sum_{p=1}^{N_R} D^2(h^c_{p,x}, h^c_{p,ref})/2.\sigma^2_C\right). \tag{3.9}$$

La figure 3.13 montre qu'une fonction de mesure sur plusieurs distributions de couleur distinctes permet *a priori* de privilégier le sujet souhaité. Néanmoins, les mouvements

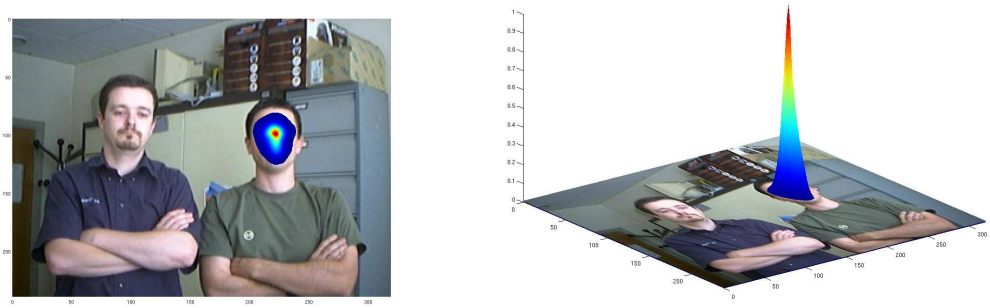


FIG. 3.13 – Exemple de vraisemblance en tout point image pour une mesure de distribution colorimétrique (deux régions d'intérêt)

réels de la cible ou les variations d'illumination induisent en général des changements d'apparence dans le flot. Pour prendre en compte ces changements, le modèle de référence est mis à jour à chaque instant comme suit [Nummiaro et al., 2002] :

$$h^c_{ref,k} = (1-\alpha).h^c_{ref,k-1} + \alpha h^c_{E[x_k]} \tag{3.10}$$

où $h^c_{E[x_k]}$ désigne la distribution liée à l'état courant estimé de la cible et $h^c_{ref,k-1}$ désigne la distribution de référence à l'instant $k-1$. Le coefficient pondère l'influence de ces deux distributions dans la mise à jour. Pour des variations rapides (resp. lentes) de l'apparence dans l'image, α doit tendre vers 1 (resp. vers 0).

3.4 Attribut forme

3.4.1 Généralités

Selon les scenarii envisagés, la nature de la cible peut être connue *a priori*, par exemple les membres corporels supérieurs dans un contexte d'interaction gestuelle. Il

est alors possible d'utiliser un attribut de forme pour la caractériser, *e.g.* en modélisant sa silhouette à l'aide d'une spline (figure **??**). Cette forme supposée rigide, est alors associée aux contours extraits dans l'image et constitue un attribut persistant. La section 3.4.2 rappelle quelques pré-requis sur l'extraction des contours dans une image couleur $z^{C_{rvb}}$. Les sections 3.4.3 et 3.4.4 décrivent deux détecteurs respectivement de formes circulaires et de visages. Différentes fonctions d'importance et de mesure, pour le seul attribut forme, sont décrites en sections 3.4.5 et 3.4.6 tandis que les sections 3.4.7 et 3.4.8 proposent des stratégies permettant d'associer cet attribut avec tout ou partie des attributs mouvement et/ou couleur vus précédemment.

3.4.2 Caractérisation des contours d'images couleurs

Intuitivement, les opérateurs de contours dédiés aux images monochromes peuvent être étendus aux images couleur représentées par leurs trois canaux notés f_1, f_2, f_3 pour respectivement R, V, B. Le gradient couleur résultant est calculé en combinant les vecteurs gradients de chaque canal f_i pris indépendamment [Herodotou et al., 1998]. Cette démarche ne permet pas de prendre en compte la corrélation entre les canaux. Ainsi, les contours de forts gradients en amplitude mais de directions opposées sur deux des trois canaux seront sous-estimés [Plataniotis et al., 2000].

Une démarche purement vectorielle constitue une alternative intéressante. L'image couleur est traitée ici comme une matrice de vecteurs aux trois canaux f_i. Platianotis *et al.* [Plataniotis et al., 2000] distinguent ici les opérateurs directionnels et les opérateurs de gradient vecteur. Pour ces derniers, l'approche proposée par Di Zenzo *et al.* [Zenzo, 1986] nous semble intéressante. Rappelons son principe.

Soient **r,v,b** trois vecteurs unitaires sur les axes R, V, B. Les dérivées directionnelles horizontale et verticale s'expriment sous la forme

$$\mathbf{u} = \frac{\partial R}{\partial x}\mathbf{r} + \frac{\partial V}{\partial x}\mathbf{v} + \frac{\partial B}{\partial x}\mathbf{b},\ \mathbf{v} = \frac{\partial R}{\partial y}\mathbf{r} + \frac{\partial V}{\partial y}\mathbf{v} + \frac{\partial B}{\partial y}\mathbf{b}.$$

Soit le gradient noté $Grad = \begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix}$ tel que

$$g_{xx} = \mathbf{u}.\mathbf{u} = \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial V}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2$$

$$g_{yy} = \mathbf{v}.\mathbf{v} = \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial V}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2$$

$$g_{xy} = \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial V}{\partial x}\frac{\partial V}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y}.$$

La direction du gradient de Di Zenzo est donnée classiquement par le vecteur propre associé à la plus grande valeur propre de $Grad$, et son amplitude est la racine carrée de cette valeur propre,soit, respectivement,

$$\theta = \frac{1}{2}\arctan\left(\ \frac{2.g_{xy}}{g_{xx}-g_{yy}}\ \right)$$

$$F(\theta) = \frac{1}{2}\{(g_{xx} + g_{yy}) + \cos 2\theta.(g_{xx} - g_{yy}) + 2.g_{xy}.\sin\theta\}.$$

Les contours sont obtenus en seuillant $\sqrt{F(\theta)}$ tandis que les dérivées images selon les directions x et y sont calculées en convoluant le vecteur $f = (f_1, f_2, f_3)$ avec les deux masques

$$\frac{\partial f_i}{\partial x} = \frac{1}{6}\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} * f_i, \quad \frac{\partial f_i}{\partial y} = \frac{1}{6}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * f_i.$$

Contrairement aux opérateurs gradient pour images monochromes étendus aux images couleur, le gradient de Di Zenzo, en prenant en compte la nature vectorielle de l'image, extrait davantage de contours couleurs. Il reste, malgré tout, très sensible aux faibles variations de texture, ainsi qu'aux bruits impulsionnels ou gaussiens.

Pour nos images couleur, nous avons utilisé cet opérateur. Cependant, pour des considérations de temps de calcul, nous avons également utilisé classiquement l'extracteur de contours proposé par Canny [Canny, 1986]. Il est appliqué sur la seule composante intensité après un changement de base qui découple les composantes intensité et chromatiques. La figure 3.15 illustre sur un exemple les opérateurs mentionnés. Pour

FIG. 3.15 – Exemples de détection de contours : (a) image originale, (b) par Sobel, (c) par Canny et (d) par Di Zenzo

plus de détails sur ces généralités, le lecteur pourra se référer à [Plataniotis et al., 2000]. Les images d'amplitude et d'orientation du gradient seront respectivement notées z^{F_c} et z^{F_o} dans la suite.

3.4.3 Détection de régions circulaires

Dans notre contexte applicatif, le but est de détecter des formes circulaires ou elliptiques correspondant à une tête et/ou une main, éventuellement les extrémités de doigts [Bretzner et al., 2002]. Nous exploitons ici le détecteur multi-échelle de *blobs* de Lindeberg *et al.* [Lindeberg, 1998] qui repose sur des invariants différentiels normalisés. Nous nous focalisons aux seules régions circulaires, le détecteur de régions elliptiques n'ayant pas encore été implémenté. Le lecteur pourra se référer à [Lindeberg, 1998] pour davantage de détails sur l'approche, notamment son extension aux régions elliptiques.

Le principe est de passer de l'espace colorimétrique RVB à l'espace Iuv de façon à découpler la composante intensité I des deux composantes chromatiques (u, v). On

rappelle la transformation[1] :

$$I = \frac{R+V+B}{3},\ u = R - V,\ v = V - B$$

Nous convoluons chacun des trois canaux $c \in \{Iuv\}$ obtenus avec un noyau gaussien $g(.;t)$ de covariance variable t. Notons $L_c(.;t)$ le résultat de cette convolution, soit $L_c(.;t) = g(.;t) * c(.)$. Pour différentes échelles t, nous sélectionnons les pixels qui localement maximisent la relation :

$$B^c_{norm} = \sum_{c \in \{Iuv\}} t^2(\partial_{xx} L_c + \partial_{yy} L_c)^2.$$

Les extrema trouvés correspondent aux *blobs* circulaires (de coordonnées x_{pos} et échelles t). La figure 3.16 montre des exemples de détection. Certaines détections peuvent être aisément filtrées, par exemple à l'aide d'une carte de probabilité couleur z^{C_p}.

FIG. 3.16 – Exemples de détection multi-échelle de régions circulaires

3.4.4 Détection de visages

Dans un contexte de suivi proximal de personne, il est intéressant d'utiliser un détecteur de visage pour définir une fonction d'importance. Il existe dans la littérature de nombreux travaux sur la détection de visages [Yang et al., 2002]. Dans cette section, nous nous intéressons plus particulièrement au détecteur multi-échelle de visages introduit par Viola *et al.* [Viola et al., 2001] puis amélioré par Lienhart *et al.* [Lienhart et al., 2002], qui offre des taux de détection excellents pour de faibles temps de calcul. La stratégie est de classifier toute sous-fenêtre de l'image en deux classes : visage ou non visage.

La méthode s'inspire de la vision humaine qui tire partie des contrastes orientés dans l'image formée sur la rétine pour interpréter la scène observée. Nous définissons des masques de Haar qui mesurent des contrastes locaux dans des directions privilégiées. Bien que n'étant pas à proprement parlé des attributs de forme, ces contrastes locaux sont classifiés dans cette catégorie du fait qu'ils sont étroitement liés à la forme caractéristique du visage humain. La différence des zones blanc/noir pour chacun des masques met en évidence les contrastes entre les yeux, les joues et le nez (figure 3.17).

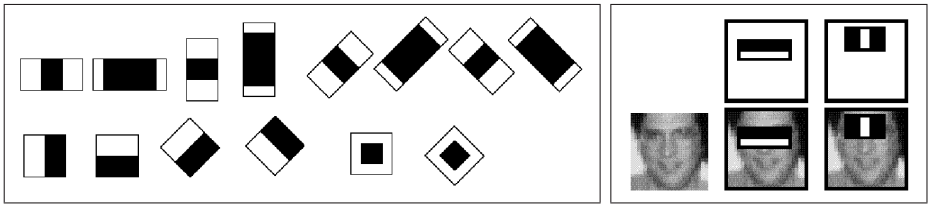


FIG. 3.17 – Masques de Haar et images d'apprentissage associées

Nous définissons alors une cascade de classifieurs qui prend en entrée toutes les sous-fenêtres de l'image en faisant varier leurs positions et leurs échelles horizontales et verticales indépendamment. Une phase d'apprentissage s'appuyant sur l'algorithme AdaBoost [Freund et al., 1995] permet de sélectionner les masques de Haar les plus discriminants pour chaque couche du classifieur. Ainsi, les premières couches filtrent un grand nombre de régions candidates pour mieux se focaliser sur les régions ambigües grâce aux couches cascadées du classifieur.

Ce détecteur de visage, illustré sur la figure 3.18, s'avère très robuste. Les images sont tirées d'une vidéo accessible à l'URL `www.laas.fr/~lbrethes`. Ce détecteur issu de la librairie OpenCV est actuellement intégré sur la plateforme mobile Rackham. Il est notamment pertinent pour démarrer une interaction entre le robot et une personne à proximité regardant la caméra, donc en situation d'interagir.

FIG. 3.18 – Exemples de détection de visages par masques de Haar

3.4.5 Fonctions d'importance associéés

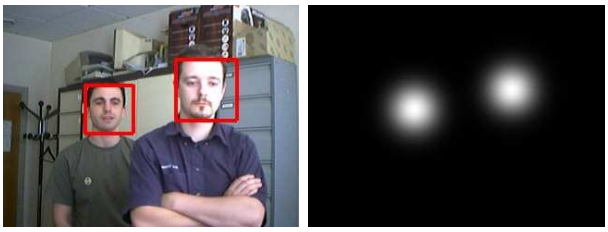
Par analogie avec (3.4) et (3.7), nous définissons la fonction d'importance $q(x_k|z_k^F)$ comme un mélange de Gaussiennes correspondant aux B régions détectées (circulaires ou visages) et de positions x_{pos_i}. Avec les notations introduites en section 3.2.2, ceci se traduit par

$$q(x_k|z_k^F) = \sum_{i=1}^{B} \delta_i \mathcal{N}(b_i, \Sigma_B), \tag{3.11}$$

avec $b_i = x_{pos_i} + \bar{x}_B$. Les entités $\bar{x}_B$ et Σ_B, estimées par apprentissage, sont respectivement la moyenne et la covariance du décalage entre la position détectée et la position

[1] En réalité, plusieurs transformations sont usuellement utilisées. Celle-ci est la plus simple.

réelle de la cible. La figure 3.19 montre sur un exemple la fonction d'importance $q(x_k|z_k^F)$ calculée en tout point image. Notons enfin que l'association de ces deux modules de dé-



(a) (b)

FIG. 3.19 – Calcul de fonction d'importance sur la forme : (a) image originale, (b) fonction d'importance sur la forme associée

tection, ou, plus globalement, de M modules de détection aboutirait à un mélange de Gaussiennes tel que

$$q(x_k|z_k^1,\ldots,z_k^M) = \sum_{j=1}^{M}\sum_{i=1}^{B_j}\delta_{i,j}\mathcal{N}(b_{i,j},\Sigma_{B_j}). \tag{3.12}$$

3.4.6 Fonctions de mesure associées

Classiquement, la vraisemblance de l'état x_k est calculée à partir de l'image de contours z^{F_c} [Isard et al., 1996c, Isard et al., 1998a]. Soient N_p points régulièrements espacés sur la spline, repérés par $x_k(j)$, $j = 1,\ldots,N_p$. Le principe est alors de rechercher sur la normale à la courbe le point de contour le plus proche de chaque point $x_k(j)$. La figure 3.20 illustre un exemple de detection de l'ensemble des points de contours sur les normales, les points les plus proches sont ensuite conservés pour la mesure.

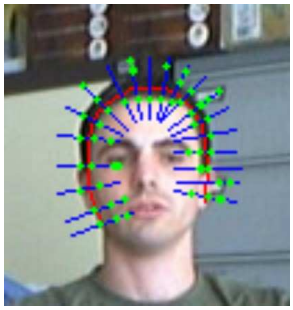


FIG. 3.20 – Recherche des points de contour selon la normale à la spline

En supposant les observations indépendantes, la densité de mesure $p(z^{F_c}|x_k)$ conditionnée sur x_k s'écrit alors

$$p(z^{F_c}|x_k) \propto \exp\left(-\frac{1}{2\sigma_{F_c}^2}\sum_{j=1}^{N_p}\phi_1(j)\right), \tag{3.13}$$

avec

$$\phi_1(j) = \begin{cases} d(j)^2 & \text{si } d(j) < \delta \\ \rho & \text{sinon} \end{cases}$$

et $d(j) = |x_k(j) - z^{F_c}(j)|$ la distance entre le point le plus proche $z^{F_c}(j)$ détecté et le point de la spline. La covariance du bruit de mesure est notée σ, δ est la distance maximale de recherche le long de la normale, enfin ρ est un coefficient de pénalité. Nous avons fixé empiriquement : $\rho = \delta^2$.

Cette mesure persistante demeure très sensible aux conditions d'éclairement *a priori* quelconques et peu discriminante pour des scènes encombrées comme illustré sur la figure 3.21. Sur cet exemple, la vraisemblance est calculée en tout point de l'image pour une échelle et orientation donnée du modèle.

Pour augmenter le pouvoir discriminant du modèle de mesure (3.13), il est judicieux de considérer l'orientation des contours dans la mesure. Tout point de contour détecté sur l'une des N_p normales à la spline doit posséder une orientation similaire à la dite normale. Dans le cas contraire, la configuration testée doit être pénalisée. Le modèle de mesure (3.13) se réécrit alors comme suit :

$$p(z^{F_c}, z^{F_o}|x_k) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N_p}\phi_2(j)\right),$$

avec

$$\phi_2(j) = \begin{cases} d(j)^2 + \Delta_\theta(j)^2 & \text{si } d(j) < \delta \\ 2\rho & \text{sinon} \end{cases}$$

et $\Delta_\theta(j) = |\theta_x(j) - \theta_{z^{F_o}}(j)|$ la différence entre l'orientation de la normale au point $x_k(j)$ et l'orientation z^{F_o} au point détecté $z^{F_c}(j)$.

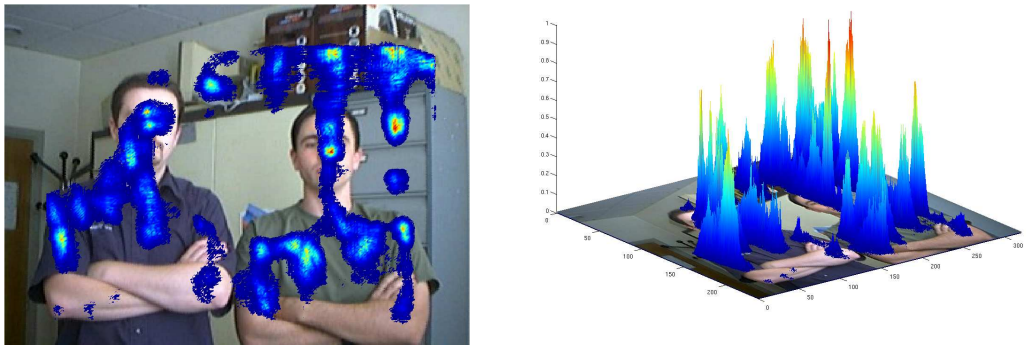


FIG. 3.21 – Exemple de vraisemblance en tout point image pour une mesure de forme (échelle et orientation fixes)

Giebel *et al.* [Giebel et al., 2004] exploitent une image de distance dans le processus d'estimation : celle-ci permet de lisser l'image de contours et reste très appropriée pour évaluer à faible coût le critère pour chacune des N particules ($N \gg 100$). Dans une image de distance, chaque pixel a une intensité fonction de la distance entre ce pixel et le point de contour le plus proche dans l'image. Pour construire cette image de distance, nous utilisons la distance de Chanfrein [Thiel, 1994]. La figure 3.22 montre un exemple et l'image de distance associée.

Le principe est alors d'attirer la cible sur les zones de faibles niveaux de l'image de distance qui soient compatibles avec la forme de la silhouette. La fonction de me-

(a) (b) (c)

FIG. 3.22 – (a) Image originale, (b) image de contours, (c) image de distance associée

sure (3.13) devient

$$p(z^{F_d}|x_k) \propto \exp\left(-\frac{1}{2\sigma_{F_d}^2}\sum_{j=1}^{N_p}\phi_1^{'}(j)\right) \tag{3.14}$$

ou $\phi_1^{'}(j)$ désigne simplement la valeur de l'intensité du pixel de l'image de distance correspondant au point $x_k(j)$ de la spline. La figure 3.23 trace la vraisemblance en tout point image pour le modèle de mesure (3.14). Comme attendu [Gavrila, 1998], nous constatons que l'exploitation d'une image de distance a pour conséquence de lisser la fonction de mesure.

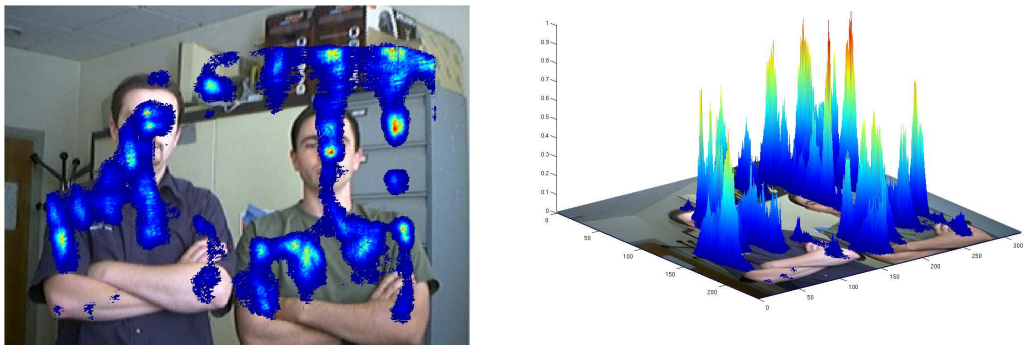


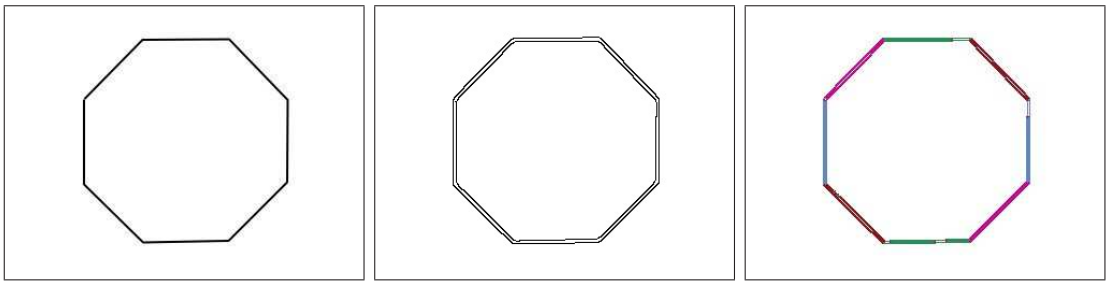
FIG. 3.23 – Exemple de vraisemblance en tout point image pour une mesure de forme basée sur l'image de distance (échelle et orientation fixes)

Pour cette nouvelle représentation, il est possible de prendre en compte l'orientation des contours dans la mesure [Gavrila, 1998]. Le principe consiste à partitionner le cercle unité en M secteurs, *i.e.* $\{[\frac{m}{M}.2\pi,\ \frac{m+1}{M}.2\pi],\ m = 0, ..., M-1\}$, et à générer M images de distance associées. Chaque point modèle $x_k(j)$ admet une orientation ψ donnée par

la normale à la spline. La fonction de mesure (3.14) devient

$$p(z^{F_d}, z^{F_o}|x_k) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N_p}\phi_2^{'}(j)\right), \tag{3.15}$$

où la valeur $\phi_2^{'}(j)$ associée au point modèle $x_k(j)$ point dans la relation (3.15) devient la valeur minimale en considérant l'ensemble des images de distance d'indices m tels que $\frac{(\psi-\varepsilon).M}{2\pi} \leq m \leq \frac{(\psi+\varepsilon).M}{2\pi}$, ε étant la tolérance donnée sur l'orientation. La figure 3.24 montre un exemple de quatre images de distance pour les directions $0°, 45°, 90°et\ 135°$.



(a) (b) (c)

FIG. 3.24 – (a) Image originale, (b) image de contours (c) images de distances orientées selon quatre directions (en fausses couleurs)

3.4.7 Combinaison avec les autres attributs dans la fonction de mesure

Combinaison avec le mouvement

Pour une caméra statique et une cible en mouvement continu ou non, nous pouvons combiner les attributs de forme et de mouvement dans la fonction de mesure. La stratégie, présentée dans [Menezes et al., 2003], est de pénaliser les pixels de contours immobiles *i.e.* admettant un vecteur vitesse apparent nul. La fonction de mesure (3.13) devient

$$p(z^{F_c}, z^{M_f}|x_k) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N_p}\phi_3(j)\right), \tag{3.16}$$

avec

$$\phi_3(j) = \begin{cases} d(j)^2 + \rho\gamma(z^{M_f}(j)) & \text{si } d(j) < \delta \\ 2\rho & \text{sinon} \end{cases}$$

et

$$\gamma(z^{M_f}(j)) = \begin{cases} 0 & \text{si } z^{M_f}(j) \neq 0 \\ 1 & \text{sinon} \end{cases},$$

$z^{M_f}(j)$ désignant l'amplitude du flot optique pour le point de contour $z^{F_c}(j)$ détecté le long de la normale à $x_k(j)$. La figure 3.25 montre la vraisemblance calculée en tout point

image pour la fonction de mesure (3.16). La personne de droite est supposée mobile et ses paramètres d'échelle et d'orientation sont supposés connus *a priori*.

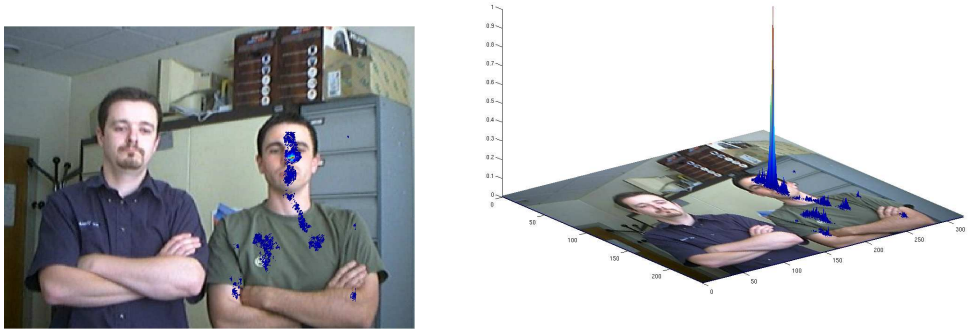


FIG. 3.25 – Exemple de vraisemblance en tout point image pour une fonction de mesure combinant forme et mouvement

Le pic de vraisemblance obtenu est très étroit tandis que les entités vraisemblables du point de vue de la forme mais statiques sont pénalisées.

Combinaison avec la couleur

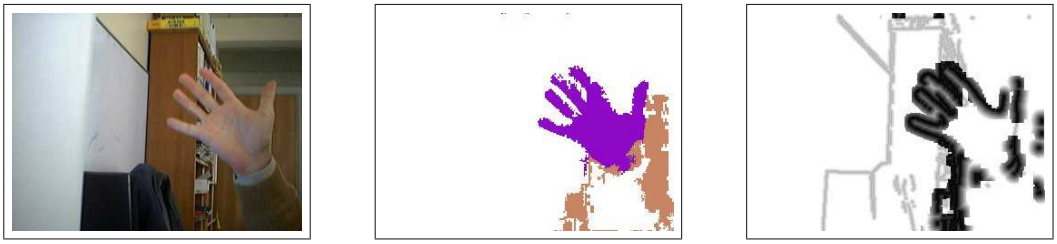
Dans [Brèthes et al., 2004b], nous avons proposé une fonction de mesure originale combinant notre segmentation régions z^{C_r} à l'attribut forme tel qu'il a été défini précédemment. Les régions peau segmentées et plus particulièrement les contours de ces régions constituent un masque z^{mask}. Les points de contour z^{F_d} situés hors de ce masque sont alors pénalisés dans l'image de distance. La nouvelle fonction de mesure $p(z^{F_d}, z^{C_r}|x_k)$ devient

$$p(z^{F_d}, z^{C_r}|x_k) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N_p}\phi(j)\right), \tag{3.17}$$

avec

$$\phi(j) = \begin{cases} d(j)^2 & \text{si } z^{mask}(j) = 1 \\ d(j)^2 + \rho & \text{si } z^{mask}(j) = 0 \end{cases}$$

où $d(j)$ est la distance mesurée dans l'image de distance et ρ désigne la pénalité apportée par le masque couleur. La figure 3.26 montre la segmentation régions et l'image de distance pondérée pour l'image de gauche. Ces exemples sont tirés d'une vidéo accessible à l'URL `www.laas.fr/~lbrethes`.



(a) (b) (c)

FIG. 3.26 – Exemple de masquage de l'image de distance à partir de la segmentation régions : (a) image originale, (b) segmentation régions, (c) image de distance pondérée

3.4.8 Fusion avec les autres attributs dans la fonction de mesure

Nous supposerons que les M sources de mesures $z_k^1, \ldots, z_k^M$ sont indépendantes, de sorte que la fonction de mesure globale s'écrit

$$p(z_k^1, \ldots, z_k^M | x_k) = \prod_{j=1}^{M} p(z_k^j | x_k). \tag{3.18}$$

Dès lors, nous pouvons définir des fonctions de mesure fusionnant plusieurs attributs i.e. : (i) forme et couleur, (ii) forme, couleur et mouvement, (iii) couleur et mouvement.

Fusion avec la couleur

Dans [Stenger et al., 2003], Stenger *et al.* proposent une fonction de mesure fusionnant forme et couleur. Celle-ci repose à la fois sur la fonction de mesure (3.13) et une carte de probabilité z^{C_p} d'appartenance à la classe peau. Les pixels sur chacune des N_p normales aux points $x_k(j)$, pour $j = 1, ..., N_p$, sont situés à l'intérieur $\mathcal{O}$ ou à l'extérieur $\mathcal{B}$ de la forme prototype. Le modèle de mesure s'écrit alors

$$p(z^{C_p} | x_k) = \prod_{o \in \mathcal{O}} p(I(o) | x_k) \prod_{b \in \mathcal{B}} p(I(b) | x_k),$$

où $I(o)$ (resp. $I(b)$) désignent les valeurs chromatiques au pixel sur la normale au point $x_k(j)$ et à l'extérieur (resp. intérieur) de la forme. Les entités $p(I(o)|x_k)$ et $p(I(b)|x_k)$ sont les distributions de probabilités peau et non peau associées. La fonction résultante est notée

$$p(z^{C_p}, z^{F_c} | x_k) = p(z^{C_p} | x_k).p(z^{F_c} | x_k).$$

Une autre démarche est d'exploiter ici notre segmentation couleur et l'image de distance sur les contours. Ainsi, soient :

- l'image de distance classique générée à partir des points de contour image et $p(z^{F_d}|x_k)$ le modèle de mesure associé ;
- une nouvelle image de distance générée à partir des contours des régions peau segmentées z^{C_r} et $p(z^{C_d}|x_k)$ le modèle de mesure associé ;

à partir de la relation (3.18), il vient alors

$$p(z^{C_d}, z^{F_c}|x_k) = p(z^{C_d}|x_k).p(z^{F_d}|x_k).$$

Contrairement à la combinaison des attributs en section 3.4.7, la démarche requiert ici la génération assez coûteuse de deux images de distance. Enfin, concernant la distribution locale de couleur et sa fonction de mesure $p(z^{C_{rvb}}|x_k)$, nous pouvons définir deux fonctions de mesure fusionnant forme et couleur :

$$p(z^{C_{rvb}}, z^{F_c}|x_k) = p(z^{C_{rvb}}|x_k).p(z^{F_c}|x_k) \tag{3.19}$$

$$p(z^{C_{rvb}}, z^{F_d}|x_k) = p(z^{C_{rvb}}|x_k).p(z^{F_d}|x_k). \tag{3.20}$$

Fusion avec couleur et mouvement

Il est aisé d'étendre ces considérations à la fusion des trois attributs, *e.g.* la figure 3.27 montre la vraisemblance en tout point image pour la fonction de mesure

$$p(z^{C_{rvb}}, z^{F_c}, z^{M_d}|x_k) = p(z^{C_{rvb}}|x_k).p(z^{F_c}|x_k).p(z^{M_d}|x_k) \tag{3.21}$$

en fixant l'échelle et l'orientation.

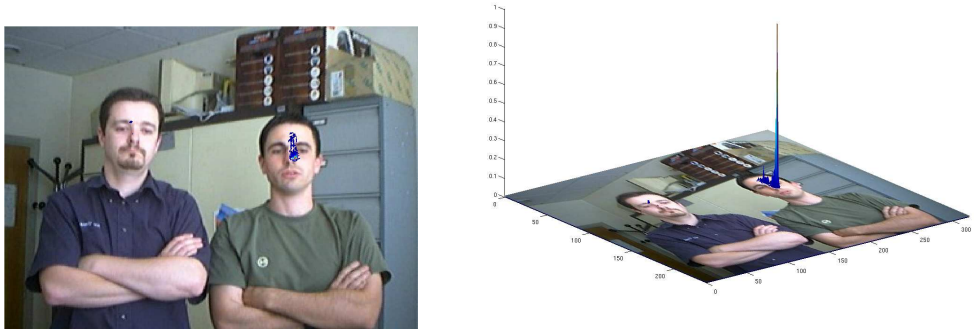


FIG. 3.27 – Exemple de vraisemblance en tout point image pour une fonction de mesure fusionnant les trois attributs

On note sur la figure un seul pic bien marqué tandis que les zones jusqu'alors privilégiées sont affectées d'une vraisemblance faible. Notons que l'on omet l'attribut forme dans la fusion, nous avons alors :

$$p(z^{C_{rvb}}, z^{M_d}|x_k) = p(z^{C_{rvb}}|x_k).p(z^{M_d}|x_k) \tag{3.22}$$

Une alternative pour associer ces trois attributs est de considérer enfin la fonction $p(z^{F_c}, z^{M_f}|x_k)$ introduite au § 3.4.7. Nous avons alors :

$$p(z^{C_{rvb}}, z^{F_c}, z^{M_f}|x_k) = p(z^{C_{rvb}}|x_k).p(z^{F_c}, z^{M_f}|x_k) \tag{3.23}$$

3.5 Évaluation globale

Plus globalement, nous avons évalué les différents attributs présentés ainsi que quelques combinaisons/fusions afin de quantifier leur pertinence. Nous avons considéré le flot vidéo d'images acquises depuis le robot durant l'exploration de l'environnement. Quelques 400 images sont extraites de ces séquences et constituent notre base pour les évaluations. Ces images sont représentatives des scènes et variations d'illumination rencontrées à l'intérieur du laboratoire, et chacune d'elles inclut au plus un visage, perçu de face ou de profil. Les situations où la personne est dos à la caméra ne sont pas envisagées dans ces évaluations car l'apparence colorimétrique de la zone d'intérêt est classiquement proche d'une teinte chair. Le processus de suivi devra néanmoins gérer ces situations en réactualisant dans le flot vidéo l'information colorimétrique de la cible (relation 3.10).

La figure 3.28 rend compte de la grande variété des conditions de prise de vue par le robot, justifiant pleinement notre démarche multicritères. Signalons enfin que l'apprentissage des distributions colorimétriques s'effectue indépendamment des images de la base, du fait qu'il repose sur la banque d'images du Compaq Cambridge Research Laboratory [Jones et al., 1998] comportant 3077 images incluant des pixels étiquetés peau et 6286 images sans pixels peau. La fonction d'importance ou de mesure relative à

FIG. 3.28 – Variété des conditions de prise de vue : exemples

un attribut ou à une combinaison est caractérisée en tout pixel image pour une échelle supposée constante et fixée *a priori*. Aucune considération spatio-temporelle n'est donc prise en compte ici, le but étant de caractériser ces fonctions indépendamment du processus d'estimation.

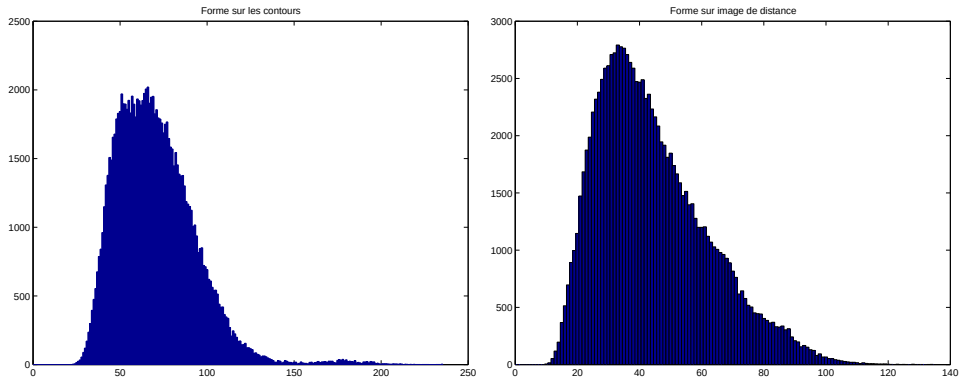
Il nous était difficile d'évaluer la totalité des fonctions de mesure ou d'importance présentées dans ce chapitre. Seul est donc évalué un panel représentatif de fonctions qui seront exploitées lors du suivi. De fait, certaines fonctions en cours d'implémentation typiquement la fonction de mesure (3.15) ou trop coûteuses en temps de calcul conceptuellement (par exemple la fonction (3.7) pour les blobs circulaires) sont omises durant ces évaluations et la suite de ces travaux.

3.5.1 Calcul des écart-types relatifs aux fonctions de mesure

Les fonctions de mesure vues précédemment incluent une constante σ dont la valeur est fixée hors-ligne. Pour cela, nous nous appuyons sur les images positives de la base, et définissons par recalage manuel les positions et échelles du template. Le principe pour estimer ces constantes est de faire varier la position du template autour de la vraie position et de calculer pour chacune de ces positions la distance (numérateur dans l'exponentielle) selon le critère (couleur, forme,...). Les distributions associées à la mesure sont supposées gaussiennes et centrées sur 0. Le tracé de la distribution des distances relevées permet d'obtenir la gaussienne incomplète associée à la mesure. Une approximation de cette gaussienne permet d'en déduire σ. Les figures 3.29.(a) à 3.29.(c) montrent les distributions obtenues pour les fonctions de mesure (3.13) et (3.14) relatives aux contours.

$$\sigma_{F_c} = 28,\ \sigma_{F_d} = 15.$$

La démarche est transposable à toutes autres fonctions de mesure vues précédemment afin d'en caractériser les constantes σ associées.



(a) (b)

FIG. 3.29 – Apprentissage des constantes σ pour les fonctions de mesure relatives à : (a) contours, (b) image de distance sur les contours

3.5.2 Fonctions de mesure

Pour évaluer l'apport de l'attribut mouvement dans les fonctions de mesure, nous avons constitué deux bases d'images avec des cibles statiques et mobiles. Chacune inclut la présence d'un visage de face ou de profil, le nombre moyen par image de vrais positifs est donc de 1.

Chaque fonction de mesure est évaluée en termes de pouvoir discriminant, précision et temps de calcul. A échelle et orientation fixe, la position est modifiée afin de déterminer les variations de la mesure dans l'image. Seuls les pics supérieurs en amplitude à la moitié du pic maximal seront considérés comme significatifs et assimilés à une détection explicite de la cible. Pour la suite, nous adoptons les notations suivantes :

- $F1$: mesure "classique" de forme sur les contours (fonction de mesure (3.13), prototype tour de tête) ;
- $F2$: mesure de forme basée sur l'image de distance des contours extraits (fonction de mesure (3.14), prototype tour de tête) ;
- $C1$: mesure sur la distribution de couleur (fonction de mesure (3.8)). Le modèle de référence est sélectionné manuellement sur la cible ;
- $F1C1$: mesure résultant de la fusion des fonctions $F1$ et $C1$ (fonction de mesure (3.19)) ;
- $F1M$: mesure résultant de la combinaison de $F1$ avec le flot optique (fonction de mesure (3.16)) ;
- $F2C1$: mesure résultant de la fusion $F2$ avec $C1$ (fonction de mesure (3.20)) ;
- $F2C2$: mesure résultant de la combinaison de $F2$ et de notre segmentation couleur (fonction de mesure (3.17)) ;
- $F1MC1$: mesure résultant de la fusion de $F1M$ et $C1$ (fonction de mesure (3.23)).

Pouvoir discriminant

Considérons tout d'abord une zone d'intérêt statique et la base d'images associée. Le graphique 3.30 montre le nombre moyen de « faux négatifs », de « faux positifs » et de « vrais positifs » par image pour les fonctions de mesure précédentes. Considérons une région d'intérêt autour de la vraie position, typiquement des positions à une distance inférieure à 40 pixels. Les vrais positifs sont alors caractérisés par des pics de vraisemblance significatifs et localisés sur la région d'intérêt. Les faux négatifs sont caractérisés par l'absence de ces pics sur la région d'intérêt. Enfin, les faux positifs sont relatifs aux pics significatifs hors de la région d'intérêt. Les fonctions de mesure $F1$ et $F2$, peu discriminantes, génèrent beaucoup de faux positifs. Les fonctions de vraisemblance combinant/fusionnant deux attributs, par exemple $F1C1, F2C1, F2C2$, limitent les faux positifs et vrais négatifs et favorisent les vrais positifs. Associer plusieurs attributs améliore logiquement le pouvoir discriminant de la fonction de mesure. Le filtre sera, en conséquence, plus robuste aux fausses mesures. On notera que la fonction $F2C1$ offre, globalement, un bon compromis.

Eu égard de l'hypothèse initiale, l'attribut mouvement n'a aucun impact sur les statistiques obtenues pour les fonctions de mesure $F1M$ et $F1MC1$. Le graphique 3.31 est relatif aux évaluations sur la base d'images pour une cible supposée mobile. Dans ce contexte, les faux positifs sont logiquement filtrés pour les fonctions $F1M$ et $F1MC1$.

Précision

A partir des vrais positifs, nous pouvons caractériser la précision de chaque fonction de mesure. Le graphique 3.32 montre l'erreur moyenne, en pixels, existant entre la vraie position et, respectivement, le pic de vraisemblance maximum ou le pic significatif qui lui est le plus proche spatialement. Dans les deux cas de figure, on note que les erreurs sont relativement faibles. Considérer les pics les plus proches n'améliore pas significativement

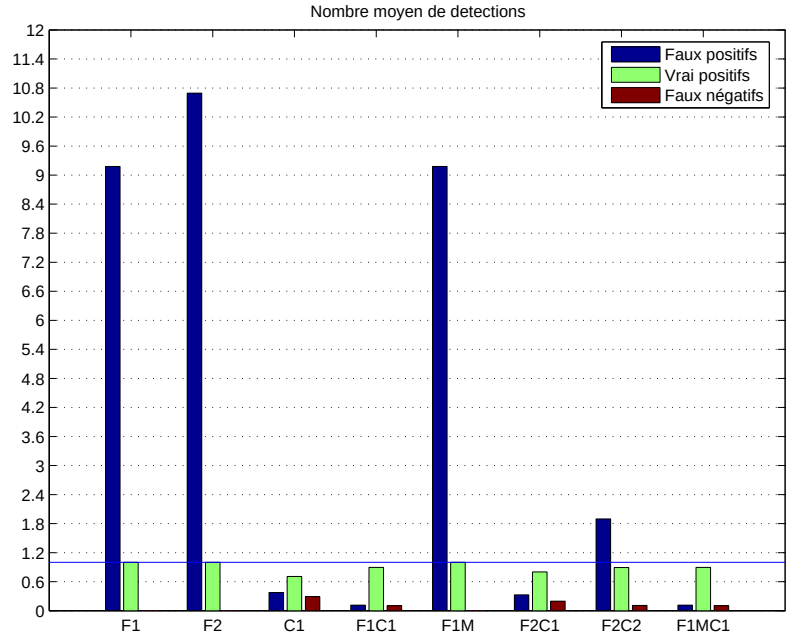


FIG. 3.30 – Nombres moyens (par image) de détections relatives aux faux/vrais positifs, et faux négatifs pour différentes fonctions de mesure (cible supposée statique)

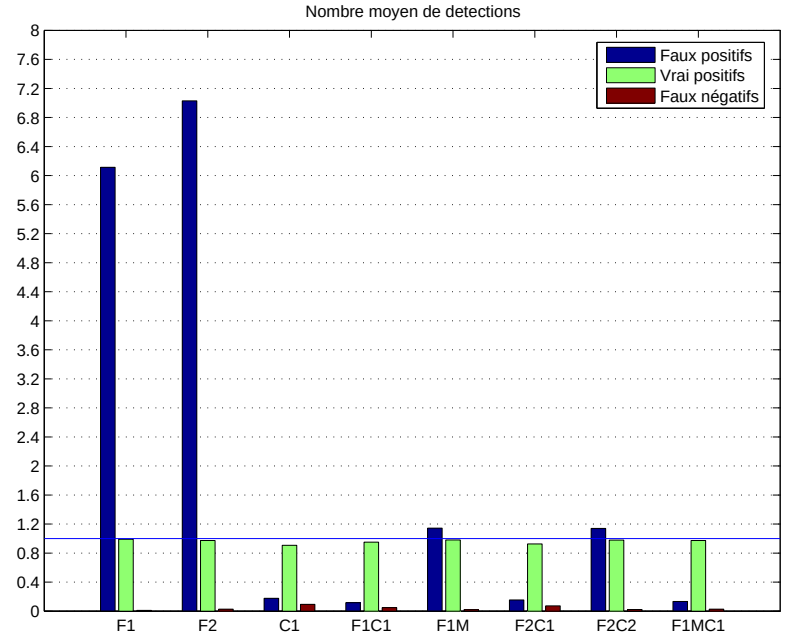


FIG. 3.31 – Nombre moyen (par image) de détections relatives aux faux/vrais positifs et faux négatifs pour différentes fonctions de mesure (cible supposée mobile)

la précision, ce qui tend à prouver que les pics de vraisemblance maximum sont souvent les plus proches des positions réelles.

Comme attendu, les fonctions de mesure basées sur la forme, i.e. $F1$ et $F2$, donnent les meilleures précisions. *A contrario*, les fonctions de mesure reposant sur la distribution locale de couleur à travers $C1, F1C1$ et $F1MC1$ donnent des précisions légèrement dégradées. Enfin, les fonctions de vraisemblance $F2C1$ et surtout $F2C2$, tout en associant les attributs forme et couleur, donnent des précisions très satisfaisantes. Les graphiques

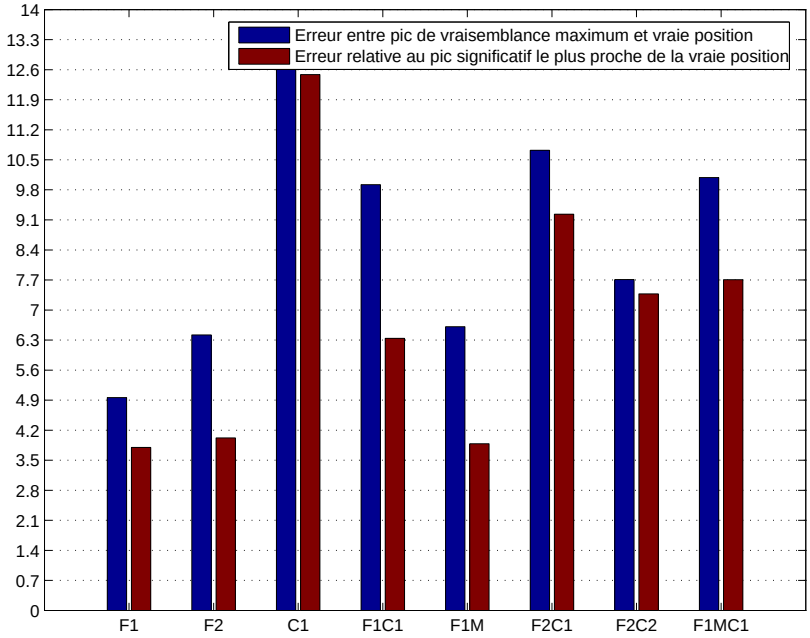


FIG. 3.32 – Erreurs moyennes en pixel entre position réelle et (1) le pic de vraisemblance maximum, (2) le pic significatif le plus proche

sur la figure 3.33 montrent la dispersion des erreurs pour chacune des fonctions de mesure. Les erreurs les plus importantes sont logiquement obtenues pour les fonctions de mesure incluant la distribution locale de couleur.

Temps de calcul

Les fonctions de mesure multi-attributs semblent améliorer les performances mais ont un coût en temps de traitement qu'il est nécessaire d'évaluer. Nos développements sont effectués sous Linux à partir des librairies OpenCV sur un PC équipé d'un processeur Pentium IV cadencé à 3 GHz.

Pour nos différentes fonctions de mesure, nous avons comparé leurs temps moyens de mise en forme : génération de l'image de distance pour la fonction de mesure $F2$, segmentation couleur pour la fonction $F2C2$, etc. Ces temps sont reportés sur le graphique 3.34. La fonction de mesure $F2$ est plus coûteuse que la fonction $F1$ car elle requiert la génération d'une image de distance sur les contours. La fonction $F2C2$ est

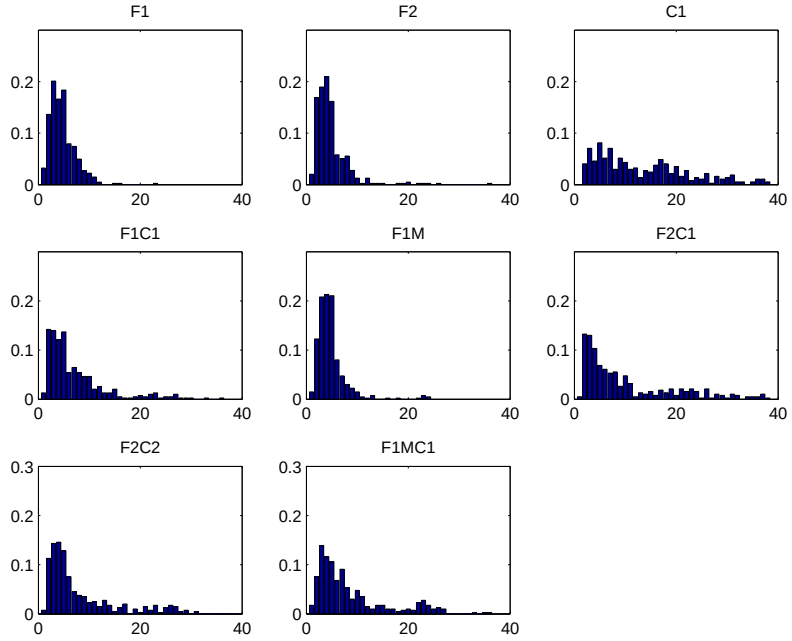


FIG. 3.33 – Histogramme des erreurs relatives à chaque fonction de mesure

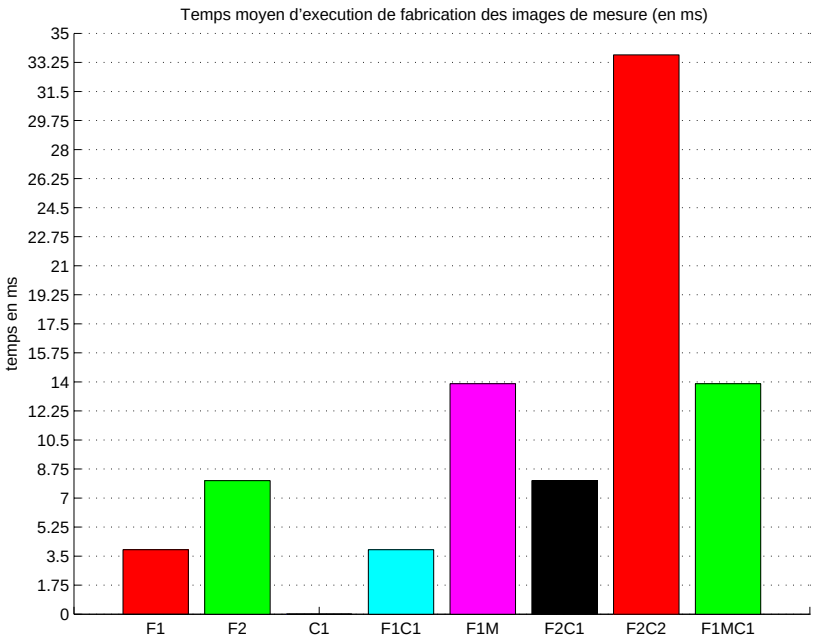


FIG. 3.34 – Coût moyen en temps de calcul associé à la mise en forme des fonctions de mesure

la plus chère car elle inclut en plus une segmentation couleur. Cependant, le temps investi dans la préparation permet un calcul plus rapide de la vraisemblance des particules comme illustré sur la figure 3.35. Pour l'attribut forme, les fonctions de mesure

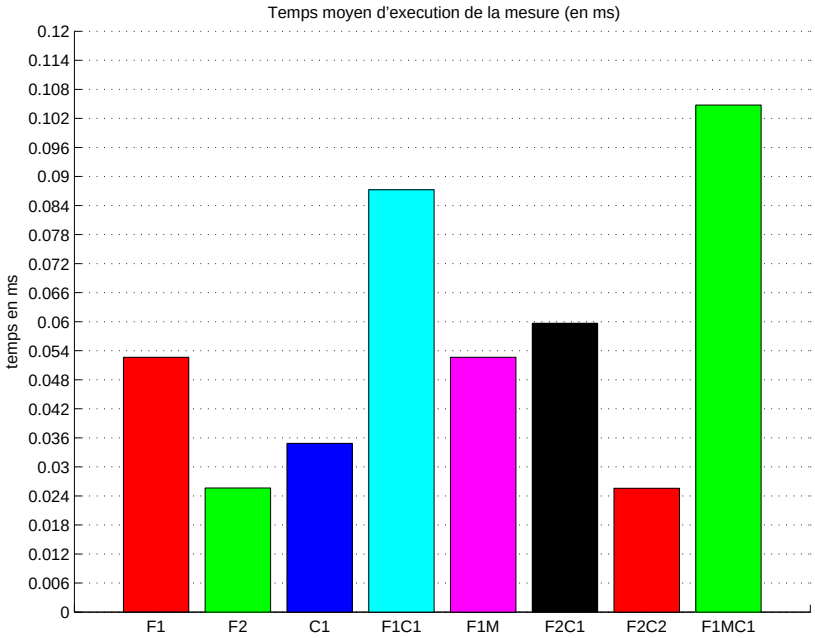


FIG. 3.35 – Coût moyen en temps de calcul associé à chaque particule

reposant sur des images de distance permettent un calcul simplifié donc peu coûteux de la vraisemblance propre à chaque particule. Le coût initial de mise en forme est donc compensé à partir d'un nombre donné de particules. Le graphique 3.36 montre le temps cumulé global en fonction du nombre de particules. Ainsi, la fonction de mesure $F2$ devient plus performante que la fonction $F1$ à partir de $N \sim 150$ particules. Parmi les fonctions associant plusieurs attributs, la fonction $F1MC1$ est globalement la plus coûteuse. Le coût de la fonction de mesure $F2C2$ croît très peu avec le nombre de particules mais reste élevé pour un faible nombre de particules. Pour $N > 100$ particules, la fonction de mesure $F2C1$ offre des temps de calcul bien inférieurs à toutes les fonctions associant deux attributs.

3.5.3 Fonctions d'importance

Les fonctions d'importance évaluées sont relatives aux différents modules de détection. Aussi, chaque image de la base de test inclut ou non la présence d'un visage. La proportion d'images positives dans la base (69%) n'obéit à aucune règle mais privilégie notre contexte d'interaction H/R où un humain est plus ou moins dans le champ de vue de la caméra. La proportion des images pour lesquelles le visage est orienté face à la caméra représente environ 65% de la base. Cette configuration, certes particulière, sera

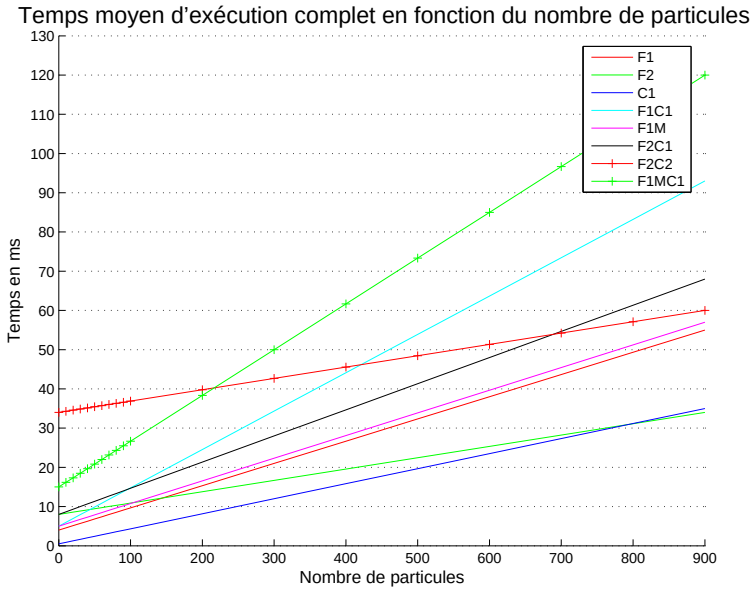


FIG. 3.36 – Temps d'exécution moyen par image en fonction du nombre de particules pour chaque fonction de mesure

abondamment exploitée dans notre contexte applicatif,[2] d'où sa prépondérance dans la base d'images.

Chacune des fonctions d'importance est évaluée en termes de pouvoir discriminant et temps de calcul. L'évaluation de la précision n'a pas vraiment de sens ici car la fonction d'importance vise à explorer adaptativement et de façon exhaustive toutes les zones « pertinentes » de l'espace d'état. L'association de plusieurs attributs revient donc ici à considérer des mélanges de Gaussiennes. Cette stratégie est assimilable à un OU logique entre ces attributs afin de détecter la cible dans un maximum de situations et donc minimiser les faux négatifs. Nous notons pour la suite :

- FD : détecteur de visage (fonction d'importance (3.11)) ;
- MD : détecteur de mouvement (fonction d'importance (3.4)) ;
- SBD : détecteur de *blobs* peau (fonction d'importance (3.7)) ;
- FMD : fonction d'importance (3.12) avec détecteurs de visage FD et de mouvement MD ;
- $FSBD$: fonction d'importance (3.12) avec détecteur de visage FD et de *blobs* peau SBD ;
- $SBMD$: fonction d'importance (3.12) avec détecteurs de *blobs* peau SBD et de mouvement MD.

[2] où le robot et son interlocuteur, supposés polis, se regardent mutuellement durant leur interaction...

Pouvoir discriminant

Comme précédemment, nous caractérisons les fonctions d'importance par le graphique 3.37 qui concerne les nombres moyens de faux/vrais positifs et de faux négatifs par image. La fonction FD délivre peu de faux positifs mais beaucoup de faux négatifs, car elle se restreint à la détection de visages quasiment fronto-parallèles. Pour la fonction SBD, les faux positifs et faux négatifs sont assez fréquents. Cette fonction détecte en effet les régions peau hors du visage (par exemple les bras ou les mains) et plus globalement les régions de l'arrière-plan proches colorimétriquement. Les non-détections sont observées lors d'une sur-exposition (robot dans un hall d'entrée) ou sous-exposition (robot dans un couloir) de la scène. Les fonctions incluant le mouvement ($MD, SBMD$) génèrent également des faux positifs et faux négatifs.

Notons enfin que les meilleures performances sont logiquement obtenues pour les fonctions FMD et $FSBD$.

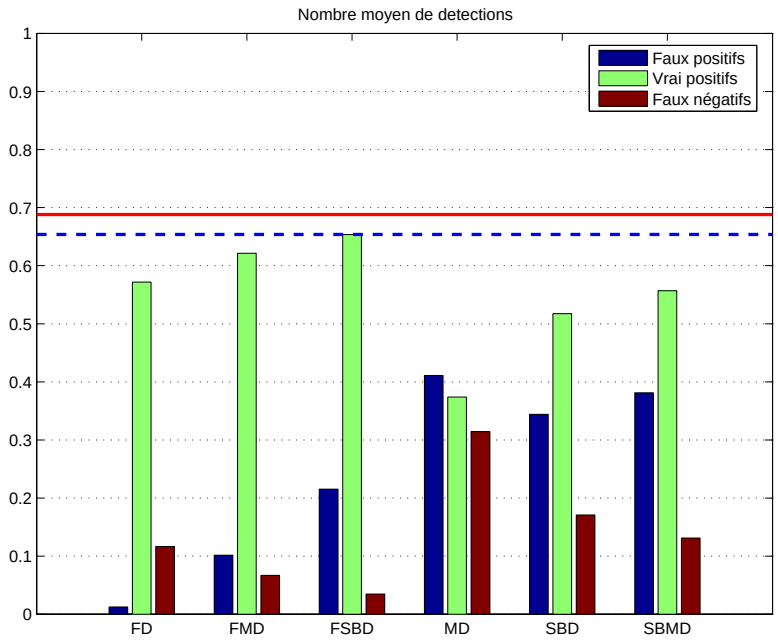


FIG. 3.37 – Nombres moyens (par image) de détections relatifs aux faux/vrais positifs, et faux négatifs pour différentes fonctions d'importance

Temps de calcul

Le graphique 3.38 montre les temps d'éxécution moyens par image des fonctions d'importance. Ces temps, exprimés en ms, restent très faibles comparativement à ceux engendrés pour les fonctions de mesure.

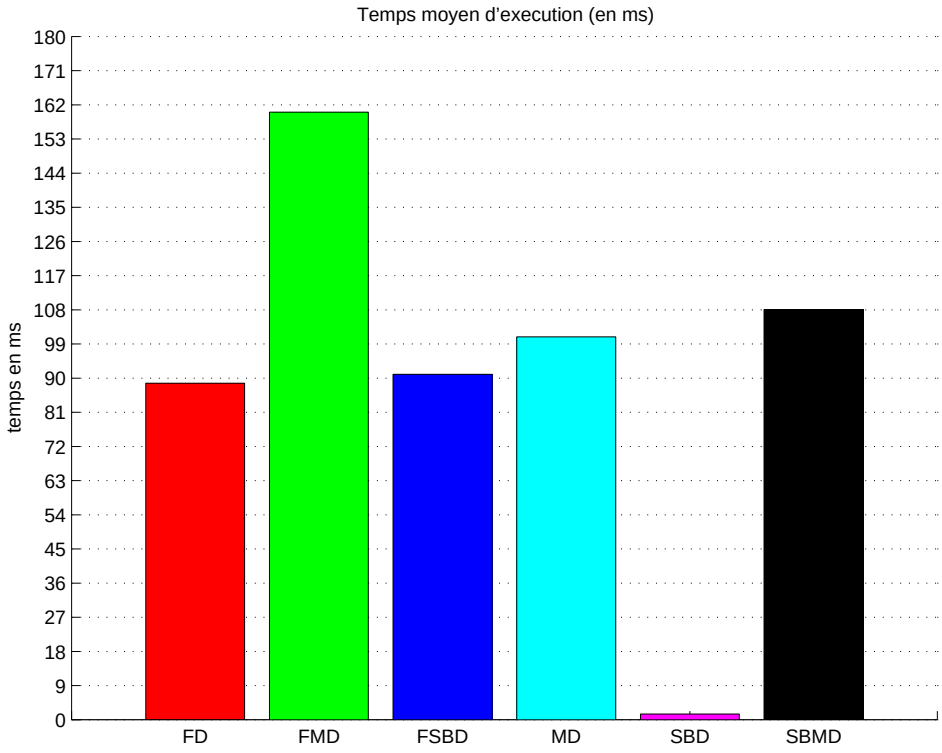


FIG. 3.38 – Temps d'éxécution moyen par image des fonctions d'importance

3.5.4 Discussion

A partir de ces évaluations, nous tirons quelques enseignements dans la perspective du chapitre suivant qui décrit des stratégies de filtrage particulaire exploitant les mesures décrites ici.

Concernant les fonctions de mesure

La fonction de mesure $F2$ basée sur l'image de distance sur les contours, peu utilisée dans un contexte particulaire (excepté [Giebel et al., 2004]), offre des performances similaires à la fonction $F1$ mais permet un gain substantiel en temps de calcul à partir de $N \sim 150$ particules.

Les fonctions associant plusieurs attributs ont logiquement un pouvoir discriminant supérieur, mais au détriment du temps de calcul. L'association de trois attributs n'apporte pas d'amélioration significative. La fonction de mesure $F2C2$, développée initialement dans [Brèthes et al., 2004b], donne de bons résultats mais son temps d'éxécution est plus élevé pour nos filtres qui comportent classiquement entre 100 et 400 particules.

Concernant la précision, il est opportun de privilégier, seul ou non, l'attribut forme dans les fonctions de mesure associées à la pondération finale des particules durant le processus d'estimation.

Au vu de ces remarques et des graphiques précédents, la fonction de mesure $F2C1$ offre un bon compromis vis-à-vis des critères mentionnés et sera donc largement utilisée par ailleurs dans nos filtres. Sans connaissance *a priori* sur la forme de la cible ou dans un contexte de surveillance (distance à la cible importante), nous privilégierons la fonction $C1$, fonction assez discriminante et peu précise pour un temps d'éxécution très faible. L'initialisation du modèle de couleur associé peut être réalisé au moyen d'un détecteur de mouvement qui indique la présence d'une cible dans l'image.

Concernant les fonctions d'importance

La fonction d'importance FD est très performante mais elle se restreint à la détection des visages quasisment fronto-parallèles et proches de la caméra ($< 3m$). Ce détecteur permet de démarrer une interaction H/R et donc de (ré)-initialiser un processus de suivi à courte et moyenne distance. En ligne, il sera associé à d'autres détecteurs, par exemple le détecteur de régions peau (fonction FSBD) qui est certes générateur de faux positifs mais peu coûteux en temps d'éxécution

Au vu du graphique 3.37, il est préférable de ne pas baser la fonction d'importance uniquement sur l'attribut mouvement. Cette remarque est corroborée par le contexte applicatif : la cible observée est par nature en mouvement intermittent. Dans ce cadre, il est souvent judicieux de placer les particules selon une fonction d'importance relative au mouvement apparent mais également selon une dynamique « centrée » sur l'état courant [Pérez et al., 2004].

Une alternative consiste à associer ici encore mouvement et *blobs* peau (fonction d'importance $SBMD$) qui donne, d'après les graphiques, des résultats corrects en termes de détection et temps d'éxécution.

3.6 Conclusion

Pour notre problématique de l'interaction Homme/Robot, nous avons présenté dans ce chapitre différentes mesures visuelles et stratégies associées afin de les combiner/fusionner dans un algorithme de filtrage particulaire. Une évaluation de ces stratégies en termes de pouvoir discriminant et temps de calcul a également été proposée en fin de chapitre.

Le formalisme du filtrage particulaire permet d'associer aisément plusieurs mesures dans les fonctions d'importance ou de mesure. Relativement peu de travaux dans la littérature exploitent cette possibilité. Nous pouvons citer ici les travaux de Isard *et al.* dans [Isard et al., 1998b] et Pérez *et al.* dans [Pérez et al., 2004]. Ces approches restent souvent confinées à un nombre restreint de primitives visuelles.

Comparativement à [Pérez et al., 2004], nous présentons ici un éventail plus large d'attributs et surtout spécifiques à notre contexte applicatif. La couleur peau, la forme caractéristique, entre autres, des membres corporels à suivre sont autant d'attributs pertinents à considérer.

Pour la fusion des attributs, Pérez *et al.* dans [Pérez et al., 2004] proposent un algorithme de filtrage hiérarchisé dans lequel deux mesures sont consommées : (1) une mesure intermittente pour éventuellement positionner efficacement les particules, (2) une mesure persistante pour pondérer ces dernières. Comme proposé dans ce chapitre, il nous semble intéressant de combiner plus largement les mesures afin d'augmenter le pouvoir discriminant des fonctions d'importance et surtout de mesure. Ainsi, la fonction d'importance pourra considérer des mesures persistantes tandis que la combinaison/fusion de mesures dans la fonction de mesure semble pertinente, notamment pour les systèmes adaptatifs qui, par définition, autorisent l'évolution des paramètres de tout ou partie des modèles de mesure. Par exemple, une distribution locale de couleur peut nécessiter une mise à jour à chaque instant image (relation 3.10), il est alors judicieux de fusionner la mesure associée avec un attribut forme. Ces deux attributs très complémentaires limitent en pratique les dérives. Plus généralement et comme illustré en section 3.4.8, la fusion de plusieurs mesures par multiplication des vraisemblances associées est assez intuitive et immédiate.

La section 3.4.7 présente, par contre, des stratégies plus fines afin de combiner plusieurs attributs dans une seule et même fonction de mesure. Dans ce cadre, nous avons proposé lors de travaux préliminaires [Brèthes et al., 2004b] un modèle de mesure original combinant couleur et forme à travers une seule image de distance afin d'évaluer à faible coût le critère relatif à chaque particule.

Les différents attributs et les stratégies de combinaison ou fusion décrites ici sont repris dans le chapitre suivant, qui présente diverses fonctionnalités de suivi par vision monoculaire couleur à partir de filtres particulaires dédiés. Ces fonctionnalités répondent à des scénarios-clés rencontrés lors de l'interaction entre un robot guide de musée et l'un des visiteurs.

Les travaux de ce chapitre ont majoritairement contribués aux publications [Menezes et al., 2003] et [Brèthes et al., 2004b] ainsi qu'à la soumission de [Brèthes et al., 2006b].

Chapitre 4

Suivi pour l'interaction Homme-Robot

Dans le chapitre 2, nous avons présenté différentes stratégies de filtrage particulaire. Le chapitre 3 a décrit un ensemble de fonctions d'importance et de mesure reposant sur divers attributs visuels. Dans ce chapitre, nous proposons des fonctionnalités de suivi associées à différentes modalités d'interaction entre l'homme et un robot, ici un robot « guide de musée ». Trois modalités sont définies ci-après. Diverses stratégies de filtrage impliquant divers attributs visuels sont proposées puis évaluées sur des séquences-types relatives à ces modalités.

4.1 Modalités d'interaction et déclinaison des fonctions visuelles associées

4.1.1 Contexte général

Rappelons que ces travaux s'inscrivent dans la problématique du robot personnel. Le défi final est de voir notre robot mobile autonome naviguer en présence de public. Durant ses tâches de planification et d'exécution de trajectoires, le robot doit prendre en compte la présence des usagers de l'environnement, par exemple en facilitant leurs déplacements. On parle ici d'interaction passive avec l'homme.

Le robot ne doit pas cependant se limiter à être un simple usager au sens où il doit pouvoir interagir avec les humains qui partagent l'environnement. On parle ici d'interaction active. Par ses déplacements, il peut chercher à interpeller les humains *via* une phase d'approche, à s'asservir sur leur déplacements, etc.

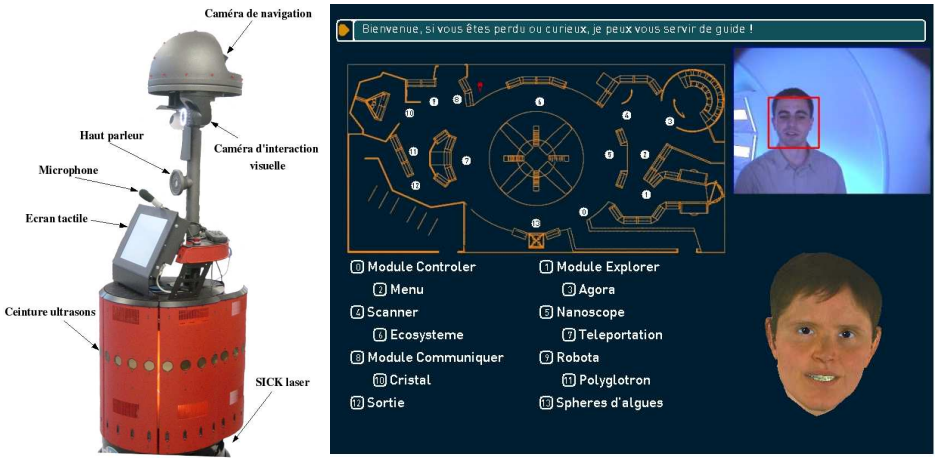
Lors de sa navigation, le robot s'appuie sur tous ses capteurs proprioceptifs et extéroceptifs (notamment caméras).

Le groupe Robotique et Intelligence Artificielle dipose de plusieurs plateformes mobiles. Nous nous focalisons sur Rackham dont les caractéristiques sont adaptées au contexte décrit précédemment. Nous le décrivons ci-après de même que les modalités d'interaction qu'il doit intégrer.

4.1.2 Rackham : le « robot guide » de la Cité de l'Espace

FIG. 4.1 – Visiteurs à la Cité de l'Espace

Dans le cadre d'une coopération entre le LAAS-CNRS et la Cité de l'Espace de Toulouse d'une part, du projet ROBEA [Bailly et al., 2005] d'autre part, un robot mobile autonome de type guide de musée a été développé. Ce robot, appelé Rackham (figure 4.2.(a)), évolue dans une exposition de la Cité de l'Espace où il guide les visiteurs qui le souhaitent vers différents stands de l'exposition. Equipé d'un télémètre laser SICK, d'une ceinture d'ultra-sons et d'une caméra numérique couleur à l'avant, il navigue de manière autonome dans l'environnement de l'exposition (figure 4.1). La figure 4.2.(a) montre le robot avec ses différents capteurs et interfaces pour l'interaction. Il est capable de se repérer dans les lieux, de planifier et d'exécuter ses déplacements y compris si l'environnement est encombré d'obstacles et de personnes. Rackham dispose aussi de diverses interfaces pour interagir avec les visiteurs. Un écran tactile leur présente le plan de l'exposition et leur propose un menu sur lequel ils sélectionnent le stand auquel ils souhaitent être conduits par le robot (figure 4.2.(b)). Sur ce même écran sont affichés un clone virtuel parlant développé par les laboratoires ICP/INPG et GRAVIR/IMAG (figure 4.2.(b) en bas à droite), ainsi que le résultat du traitement des images issues d'une caméra couleur dédiée à l'interaction (figure 4.2.(b) en haut à droite).



(a) (b)

FIG. 4.2 – (a) le robot Rackham avec ses capteurs et interfaces, (b) informations présentées sur son écran tactile

Grâce à une synthèse vocale, le clone peut parler aux visiteurs pour les interpeller ou les renseigner sur l'exposition. Un micro permet en retour au robot d'enregistrer le visiteur et de le comprendre par le biais d'un module de reconnaissance vocale.

Dans ces travaux, l'interaction repose sur les capteurs visuels embarqués car ils permettent une interaction à distance et diversifiée de par la richesse du signal vidéo délivré. Nous privilégions la caméra située à l'arrière du robot. Cette caméra, à courte focale, est montée sur une platine site-azimut. Elle est exploitée pour observer tout visiteur en situation d'interaction.

4.1.3 Description d'un scénario type

Définissons un scénario-type à partir duquel le robot Rackham est censé guider les visiteurs dans un musée.

- **Phase** #1 **:** Á l'état inactif, le robot est *a priori* immobile et guette à distance l'arrivée d'un visiteur. Le robot interpelle alors les visiteurs par son synthétiseur vocal. Pour appuyer sa volonté d'interagir, il peut effectuer quelques mètres en direction de tout visiteur détecté et suivi visuellement. Cette modalité d'interaction, éventuellement passive, est relative à une première fonction de **suivi "éloigné" ou surveillance**.
- **Phase** #2 **:** Le visiteur intéressé peut alors communiquer avec le robot par écran tactile, reconnaissance vocale ou gestuelle (figure 4.1). Indépendamment de la reconnaissance gestuelle abordée au chapitre suivant, cette interaction proximale met en évidence une seconde fonction de **suivi proximal**.
- **Phase** #3 **:** Le robot dispose d'un plan appris automatiquement durant l'exploration préalable du site. Cette carte, illustrée sur la figure 4.2.(b), permet (1) au visiteur de choisir le stand à visiter *via* l'écran tactile, (2) au robot de se déplacer au lieu désiré après avoir préalablement planifié sa trajectoire. Durant l'exécution de cette trajectoire, le robot se localise à chaque instant grâce à ses capteurs embarqués. Alors qu'il s'avance vers le stand désiré, il s'assure visuellement que le visiteur le suit. Les mouvements relatifs robot/visiteur, le champ de vue limité de la caméra, voire des occultations ponctuelles peuvent cependant aboutir à la perte du visiteur guidé. Celui-ci devra alors de lui-même se replacer dans le champ de vue ou presser un bouton de confirmation sur l'écran. Cette modalité est relative à une troisième fonction de **suivi proche**, qui devra inclure notamment une étape de réinitialisation automatique.

Revenons en détail sur ces différentes modalités complémentaires d'interaction, et déclinons les fonctions de suivi, ou « trackers » associées.

4.1.4 Modalités d'interaction-Fonctions de suivi associées

Le scénario générique présenté ci-dessus a mis en exergue les trois modalités d'interaction et fonctions de suivi suivantes :

1. **suivi "éloigné" ou surveillance :** robot en attente d'interaction et surveillant des visiteurs potentiels ($distance\ H/R > 3m$) ;
2. **suivi proximal :** robot en interaction proximale avec un visiteur (robot à l'arrêt, $distance\ H/R < 1m$) ;

3. **suivi proche :** robot guidant un visiteur dans le lieu de son choix (robot en mouvement, *distance* H/R entre $1m$ et $3m$ environ).

Le **suivi proximal** est relatif à la phase #2 du scénario précédent où le robot est nécessairement immobile. Son interlocuteur interagit avec lui *via* l'écran tactile, le micro pour la commande vocale et/ou la caméra arrière pour la commande gestuelle. Il est situé à moins d'un mètre, et tend à rester immobile et de face. La figure 4.3.(a) illustre un exemple de situation correspondant à cette modalité.

(a) (b) (c)

FIG. 4.3 – Situations d'interaction : (a) proximale, (b) à courte distance, (c) à moyenne distance (ou surveillance)

Le **suivi proche** est lié à la phase #3 du scénario précédent. Le robot est donc mobile et le visiteur est censé le suivre à une distance relativement courte (figure 4.3.(b)). Dans ce cadre, la distance relative H/R est naturellement comprise entre $1m$ et $3m$. La '« cible » peut temporairement être absente du flot vidéo de par : (1) les occultations par une autre personne - inévitable dans un musée -, (2) les mouvements conjoints du robot et de la cible. Celle-ci sort du champ de vue de la caméra, il est alors nécessaire de réinitialiser automatiquement le suivi de manière adéquate. Certes, les commandes en vitesse appliquées au robot permettent d'anticiper et d'orienter convenablement la caméra. Il apparaît cependant, que tous ces mouvements simultanés (robot, caméra, cible) induisent des mouvements apparents de la cible quelconques et sans réelle cohérence temporelle.

Notons enfin que dans cette modalité, l'attention du visiteur n'est pas en permanence focalisée sur le robot comme précédemment. Il détourne naturellement le regard, éventuellement pivote sur lui-même durant le processus de guidage.

Concernant le **suivi "éloigné" ou surveillance**, le robot est immobile et observe l'environnement (phase #1 du scénario). En attente d'une éventuelle interaction, il surveille les déplacements des personnes situées à $3m$ et au delà (figure 4.3.(c)). Pour ces distances, les risques d'occultation sont plus importants que pour la phase #3. Notons que cette situation d'attente peut ou non (au choix du superviseur...) être compatible

avec la modalité #2 du scénario : une personne plus proche du robot peut lancer une interaction proximale *via* l'activation des interfaces.

D'après le scénario décrit, le robot peut éventuellement (au choix du superviseur) se déplacer de quelques mètres en direction d'un visiteur. Ce comportement du robot n'implique pas de fonction visuelle spécifique en raison de la nature supposée brève et courte du déplacement induit. Les séquences d'images traitées et présentées ici sont toutes issues acquises depuis la caméra arrière et pour une focale fixe. A terme, nous exploiterons naturellement la caméra située en haut du mât - au dessus des têtes - pour cette modalité. La caméra utilisée n'étant pas dotée de zoom, les acquisitions d'images relatives à cette modalité sont effectuées actuellement à focale fixe.

La figure 4.4 schématise le découpage de l'espace d'interaction du robot selon nos trois modalités. Ces zones sont déterminées empiriquement à partir de l'observation du comportement habituel des personnes interagissant avec le robot.

Ces différentes modalités étant ébauchées, il nous faut alors proposer des fonctions de suivi visuel adaptées. Pour ce faire, nous avons à disposition des stratégies de filtrage (chapitre 2) et des fonctions d'importance et de mesure reposant sur divers attributs visuels (chapitre 3).

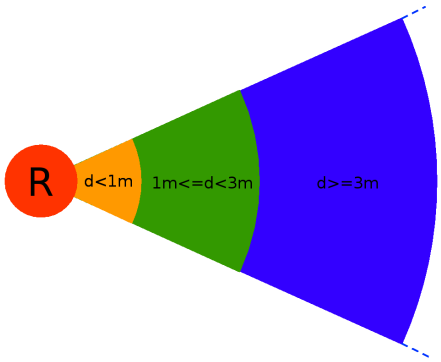


FIG. 4.4 – Distances relatives H/R pour nos modalités

4.1.5 Protocole d'évaluation

Les fonctions de suivi proximal, suivi proche et surveillance sont détaillées et évaluées, en termes de stratégies de filtrage et de mesures, dans les sections 4.2 à 4.4. Les choix seront argumentés à partir des évaluations effectuées sur des séquences tests acquises depuis le robot. Les contextes de prise de vue varient et sont caractéristiques des situations rencontrées par notre robot durant l'exploration de l'environnement : couloirs *a priori* sur- ou sous-éclairés, espaces ouverts éventuellement encombrés, occultations éventuelles de la cible,... À notre connaissance, peu de travaux proposent, certes dans un cadre applicatif très précis, des évaluations et comparaisons aussi poussées entre les nombreuses stratégies de filtrage particulaire pour la fusion de données.

Chaque séquence acquise est dépouillée afin de caractériser le vecteur d'état pour chaque image et ainsi constituer une « vérité terrain » pour les évaluations ultérieures. La démarche est soit manuelle, soit semi-automatique. Elle consiste à « lancer » un filtre admettant un nombre considérable de particules, et de le réinitialiser manuellement lors d'éventuels décrochages durant le suivi.

Sur l'ensemble des séquences, nous évaluons alors les stratégies de filtrage envisagées pour chaque modalité, selon les critères définis par Torma et Szepesvári dans [Torma et al., 2003], à savoir :

– **la précision (en *pixel*) :** quantifiée par l'erreur entre la position estimée du

template et sa vraie position à chaque instant ;
- **le taux d'échec (en %) :** quantifié par le nombre de décrochages observés, chacun étant notifié lorsque la distance précédente est supérieure à un seuil préalablement fixé.

Enfin, le **temps de traitement (en** *ms***)** sera également évalué car il représente un critère essentiel pour nos plateformes embarquant des ressources CPU limitées.

Le nombre N de particules influence grandement les performances d'un filtre particulaire. Aussi, chaque stratégie est évaluée sur chaque séquence pour $10 < N < 900$. Enfin, le caractère aléatoire du filtrage particulaire ne permettant pas de baser son évaluation sur une seule de ses réalisations, une étude statistique du comportement moyen du filtre est effectuée. Ainsi, les erreurs, taux d'échec et temps moyens sont calculés pour chaque stratégie à partir de 20 réalisations appliquées sur chaque séquence.

Signalons enfin que l'appel spécifique à une fonction visuelle, voire la commutation d'une fonction à une autre sont liées à l'état courant global du robot et au contexte environnemental. Étant gérées au niveau du superviseur, elles sortent par conséquent du cadre de cette thèse et ne sont pas étudiées ici. Des heuristiques reposant sur l'échelle du *template* dans l'image sont également envisageables en pratique.

4.2 Suivi pour l'interaction proximale

4.2.1 Considérations génerales

Pour cette modalité[1], le vecteur d'état x_k du filtre doit contenir quatre composantes liées respectivement à la position (u_k, v_k), l'orientation θ_k et l'échelle s_k du *template* (tour de tête) où k indexe ici le temps image dans le flot vidéo. Certes, les composantes échelle et orientation varient assez peu pour cette modalité mais elles permettent un recalage fin du *template*. L'orientation sera à terme exploitée afin d'interagir par des signes de la tête. Concernant la dynamique de la cible, nous avons logiquement opté ici pour une marche aléatoire (ou *random walk*) qui nous semble caractériser le mieux les mouvements apparents de la cible pour cette modalité. Dans ce modèle de dynamique, les composantes du vecteur d'état sont supposées évoluer indépendamment suivant des gaussiennes centrées sur les valeurs estimées de l'état à l'instant précédent. Le vecteur d'état du filtre est alors defini par :

$$x_k = [u_k, v_k, \theta_k, s_k]'$$

Le *tracker* associé à cette modalité est défini par ses fonctions de mesure et d'importance ainsi que sa stratégie de filtrage. Celles-ci sont déclinées ci-après puis évaluées.

4.2.2 Fonctions de mesure envisagées

La figure 4.8 montre quelques images de séquences traitées pour cette modalité. À cette distance, la fonction de mesure repose logiquement sur la silhouette (grossière) caractéristique de la cible représentée ici par une spline rigide (figure 4.5).

[1] *a priori* la plus simple

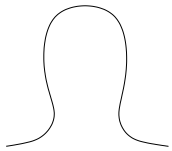


FIG. 4.5 – Silhouette de la cible

Il est opportun de ne pas considérer l'attribut forme seul dans la fonction de mesure, notamment en présence de scènes encombrées. Nous privilégions *a priori* les fonctions de mesure $F2C1$, qui fusionne forme et distribution de couleur, et $F1M$ qui combine forme et mouvement (§ 3.5). La mesure $F1M$ semble tout indiquée ici car le robot est à l'arrêt pour cette modalité ; l'attribut mouvement est donc facilement exploitable. Les mouvements réels de la cible sont certes faibles mais induisent des mouvements apparents significatifs car la distance relative H/R est faible. Le flot optique permet de filtrer les régions[2] de l'arrière-plan vraisemblables du point de vue de la forme. Des évaluations sur des images acquises pour cette modalité ont montré que les fonctions $F1M$ et $F2C1$ donnent des performances similaires en termes de précision et taux d'échec ,*i.e.* pertes de la cible, et ce quelle que soit la stratégie de filtrage. La mesure $F2C1$ reste cependant plus coûteuse en temps de calcul car le visage occupe *a priori* une grosse proportion de l'image ce qui implique de calculer la distribution de couleur dans une région de taille conséquente. Les figures 4.6 illustrent ces propos pour un filtre de type SIR et une fonction d'importance relative à la dynamique, pour des séquences traitées représentatives de la modalité considérée.

4.2.3 Fonctions d'importance et stratégies de filtrage envisagées

Plusieurs fonctions d'importance restent cependant envisageables pour cette modalité :

1. une fonction d'importance relative à la seule dynamique notée ici q_{FID} (§2.2.1) :

$$q_{FID}(x_k|x_{k-1}, z_k) = \mathcal{N}((x_k|x_{k-1}), \Sigma)$$

   avec $\Sigma = diag(\sigma_u^2, \sigma_v^2, \sigma_\theta^2, \sigma_s^2)$ une matrice diagonale définie par les variances relatives aux variations des composantes de l'état.

2. une fonction d'importance relative aux seules mesures sur la position et notée ici $q(u_k, v_k|z_k)$. La fonction d'importance globale q_{FIM} s'écrit (§2.2.2) :

$$q_{FIM}(x_k|x_{k-1}, z_k) = q(u_k, v_k|z_k).\mathcal{N}((\theta_k, s_k|\theta_{k-1}, s_{k-1}), \Sigma_d)$$

   avec $\Sigma_d = diag(\sigma_\theta^2, \sigma_s^2)$ une matrice diagonale définie par les variances relatives aux variations des composantes orientation et échelle du vecteur d'état.

3. une fonction d'importance q_{FIDM} *a priori* plus performente que les deux précédentes, relative à la fois à la dynamique et aux mesures (§A). Nous considérons ici, un mélange de fonctions d'importances qui échantillonnent un pourcentage α de particules selon $q(x_k|z_k)$, un pourcentage β selon la loi $q(x_0)$ (prior) et le reste $(1 - \alpha - \beta)$ selon la dynamique du système $p(x_k|x_{k-1})$:

$$\begin{aligned} q_{FIDM}(x_k|x_{k-1}, z_k) &= \alpha\left[q(u_k, v_k|z_k)\mathcal{N}((\theta_k, s_k|\theta_{k-1}, s_{k-1}), \Sigma_d)\right] + \beta q_0(x_k) + \\ &\quad (1 - \alpha - \beta)\mathcal{N}((x_k|x_{k-1}), \Sigma) \end{aligned}$$

[2] *a priori* peu nombreuses dans ce contexte

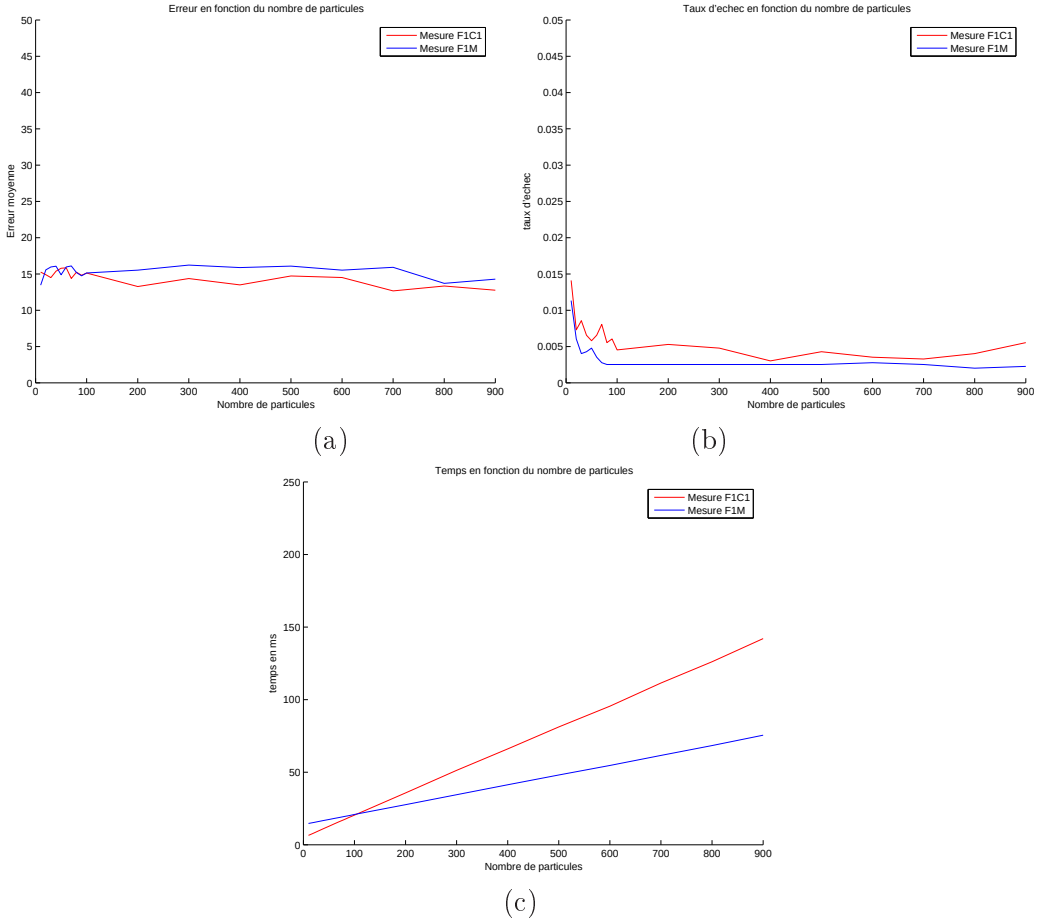


FIG. 4.6 – Erreur (a), taux d'échec (b), temps de calcul (c) *vs* nombre de particules pour les mesures $F1M$ et $F2C1$

avec $\Sigma = diag(\sigma_u^2, \sigma_v^2, \sigma_\theta^2, \sigma_s^2)$, $\Sigma_d = diag(\sigma_\theta^2, \sigma_s^2)$ deux matrices diagonales. Le mélange évalué est composé d'une petite proportion de particules tirées selon le prior $\beta = 0.1$. Les autres particules se répartissent entre les deux autres parties de la fonction d'importance avec $\alpha = 0.3$.

Pour cette modalité, nous fixons : $(\sigma_u, \sigma_v, \sigma_v, \sigma_s) = (8, 4, 10^{-4}, 5.10^{-3})$

Deux stratégies de filtrage sont envisagées et évaluées pour ces fonctions d'importance, à savoir : (1) SIR et AUXILIARY pour les fonctions de type FID[3], (2) MSIR et RBSS pour les fonctions de types FIM ou $FIDM$. Dans le cas des fonctions de type $FIDM$, les stratégies MSIR et RBSS sont combinées à la stratégie SIR pour permettre

[3]Notons que le filtre AUXILIARY est classé dans les évaluations des stratégies FID - fonction d'importance dépendant de la dynamique - du fait qu'après rééchantillonnage auxiliaire, les particules sont propagées au moyen de la dynamique *a priori* du système. Il n'en demeure pas moins que cet algorithme peut être vu globalement comme étant basé sur une fonction d'importance tenant compte à la fois de la dynamique et de la mesure

la prise en compte de la dynamique dans la fonction d'importance.

Pour les stratégies FIM et $FIDM$, nous optons sans surprise (§ 3.5) pour une fonction d'importance $q(u_k, v_k|z_k)$, dans q_{FIM} et q_{FIDM}, associant deux détecteurs respectivement de visage et de *blobs* peau. A courte distance, le détecteur de visages s'avère en effet efficace et très rapide... mais limité aux seuls visages orientés face à la caméra[4]. Nous ajoutons donc le détecteur de blobs peau, détecteur plus génerique pour un temps additionnel faible (§ 3.5). Ce dernier reste très sensible aux conditions d'illumination et aux régions parasites de couleur peau. Les évaluations du § 3.5 ont néanmoins montré que cette association dans la fonction d'importance, notée $FSBD$, minimise les vrai négatifs. La figure 4.7 montre quelques exemples de détections sur des images représentatives de cette modalité. Signalons que le détecteur de mouvement (§ 3.2.2) est inadapté ici car la précision et les temps de calcul (dûs à l'échelle) sont prohibitifs dans ce contexte.

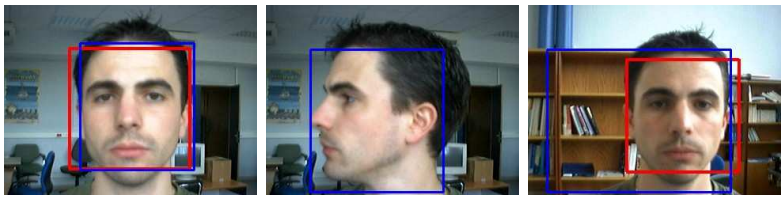


FIG. 4.7 – Exemples de détections conjointes de visages (en rouge) et blobs peau (en bleu)

À partir de ces considérations, six stratégies de filtrage, labellisées dans le tableau 4.1 sont donc identifiées et seront évaluées. Elles diffèrent par le choix de l'algorithme de filtrage mis en œuvre pour estimer l'état de la cible et par le type de fonction d'importance. Notons qu'en l'absence de détection pour les stratégies FIM, le suivi est alors assuré par un algorithme de CONDENSATION classique. En effet, la nature intermittente des mesures oblige d'envisager une alternative pour le suivi basé sur FIM. En cas de non détection, aucun échec n'est donc comptabilisé pour les stratégies FIM qui propagent alors ponctuellement les particules selon la dynamique. Ce choix favorise certes cette stratégie mais reste la seule solution pour assurer un suivi et permettre une évaluation.

4.2.4 Evaluation des stratégies de filtrage envisagées

Dans cette section, nous présentons une évaluation de ces stratégies afin de déterminer les plus adaptées à notre modalité d'interaction proximale. Les évaluations portent sur 20 séquences (20×20 réalisations au total) représentatives de la grande diversité des lieux et conditions de prises de vues rencontrés par le robot dans le laboratoire. Pour cette modalité, nous considérons entre autres des séquences de ~ 200 images acquises dans des couloirs (sous ou sur-éclairés) ainsi que dans des espaces ouverts présentant des

[4] cette situation reste néanmoins très fréquente et naturelle lors d'une interaction proximale...

Nom de stratégie	Type de filtre particulaire	Fonction d'importance
$FID1$	SIR	$q_{FID}(x_k\|x_{k-1}, z_k)$
$FID2$	AUXILIARY	
$FIM1$	MSIR	$q_{FIM}(x_k\|x_{k-1}, z_k)$
$FIM2$	RBSS	
$FIDM1$	MSIR	$q_{FIDM}(x_k\|x_{k-1}, z_k)$
$FIDM2$	RBSS	

TAB. 4.1 – Stratégies envisagées pour l'interaction proximale

arrière-plans *a priori* encombrés et donc générateurs de fausses mesures. La figure 4.8 montre quelques images issues de ces séquences.

FIG. 4.8 – Images de séquences acquises en interaction proximale (modalité #1)

La figure 4.9 illustre, pour les filtres $FID1$ et $FIM1$, une réalisation sur une séquence donnée (scène peu encombrée). la particule en rouge est la moyenne *a posteriori* estimée par le filtre *via* l'approximation :

$$[\hat{x}_{k|k}]_{MMSE} = \sum_{i=1}^{N} w_k^{(i)}.x_k^{(i)}$$

FIG. 4.9 – Un exemple de réalisation pour une séquence donnée (scène peu encombrée) avec filtres $FID1$ en haut et $FIM1$ en bas (modalité #1)

La figure 4.10 compare, pour les différentes stratégies, les erreurs et taux d'échec obtenus en fonction du nombre de particules sur des séquences de scènes encombrées.

On note que les erreurs relatives aux stratégies FID sont plus faibles. Les taux d'échec, similaires quelle que soit la stratégie, sont de l'ordre de 2% à partir de $N = 100$ particules.

Les filtres sont donc peu perturbés par l'encombrement de la scène car celle-ci reste majoritairement occultée par la cible. La présence éventuelle de régions parasites « peau » à l'arrière-plan induit une légère diminution de la robustesse et de la précision pour les stratégies FIM. Les stratégies FID et $FIDM$ ont des comportements similaires.

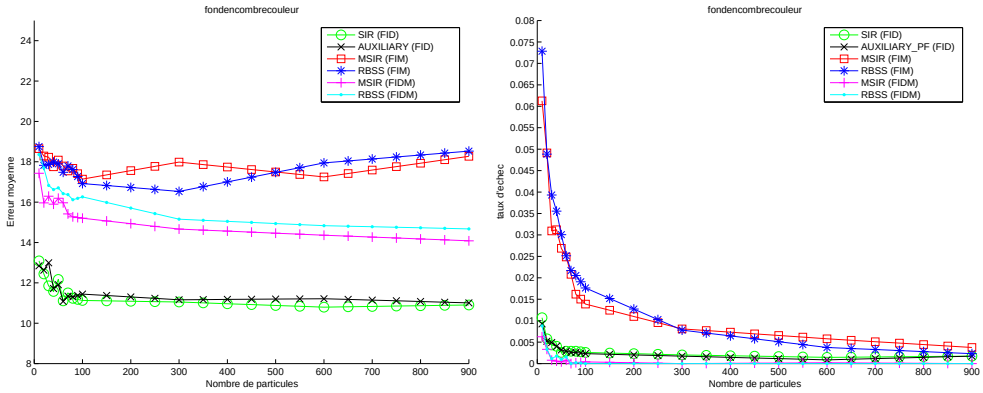


FIG. 4.10 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences de scènes encombrées pour chaque stratégie (modalité #1)

La figure 4.11 compare, pour les différentes stratégies, les erreurs et taux d'échec obtenus en fonction du nombre de particules sur des séquences incluant des forts changements d'illumination, typiquement dans un couloir.

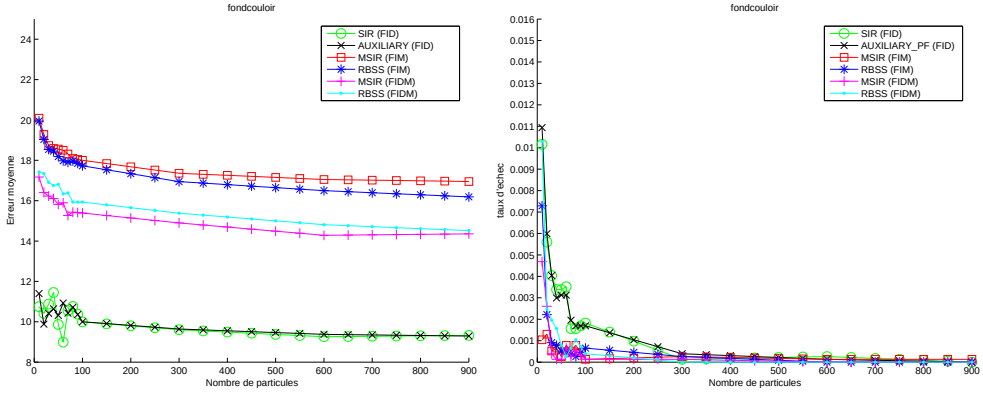


FIG. 4.11 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences de scènes avec illuminations changeantes pour chaque stratégie (modalité #1)

Les précisions, toujours favorables aux stratégies FID, sont globalement peu altérées par les variations d'illumination. Le visage reste bien contrasté à cette faible distance

($\sim 1m$). La fonction de mesure sur la forme est alors peu sensible aux variations d'illumination et permet le recalage en final du *template*. Pour les stratégies FIM et $FIDM$, le détecteur de *blobs* « peau », parfois inactif, est suppléé par le détecteur de visages moins sensibles aux changements d'illumination.

4.2.5 Discussion

Pour cette modalité d'interaction proximale, nous optons au final pour une stratégie FID car : (1) les précisions obtenues sont meilleures, (2) les très faibles taux d'échec justifient assez peu une étape de détection, (3) les temps de traitement sont plus faibles comme illustrés sur la figure 4.12.

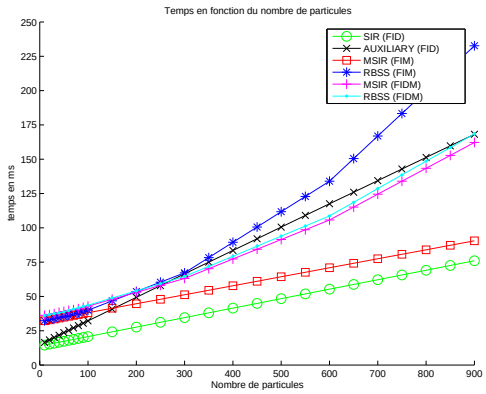


FIG. 4.12 – Temps moyen de traitement *vs* nombre de particules pour chaque stratégie (modalité #1)

Parmi les stratégies FID, la stratégie AUXILIARY n'améliore pas sensiblement les performances. De part la dynamique du système, les rééchantillonnages sont ici peu efficaces et coûteux en temps de traitement. Nous privilégions donc la CONDENSATION classique pour cette modalité ($FID1$).

4.3 Suivi pour l'interaction à mi-distance

4.3.1 Considérations générales

Dans cette modalité, le robot a pour mission de guider son interlocuteur vers un lieu prédéfini. Le robot exploite le flot vidéo de sa caméra embarquée pour adapter sa vitesse et vérifier la présence de son locuteur durant la mission. Le vecteur d'état x_k doit contenir les seules composantes position (u_k, v_k) et échelle s_k où k indexe le temps image dans le flot vidéo. Les mouvements réels de la cible sont très variables même si leur cap est donné par le robot. Les mouvements apparents associés sont d'autant plus difficiles à caractériser qu'ils résultent de la projection de mouvements relatifs entre la cible et la caméra, laquelle est montée sur platine et embarquée sur le robot mobile. Nous optons donc comme précédemment pour un modèle de type marche aléatoire qui

permet de prendre en compte toutes ces incertitudes mais avec un bruit de dynamique inchangé *i.e.* $(\sigma_u, \sigma_v, \sigma_s) = (8, 4, 5.10^{-3})$. Nous posons pour la suite :

$$x_k = [u_k, v_k, s_k]'$$

4.3.2 Fonctions de mesure envisagées

FIG. 4.13 – Régions d'intérêt caractérisées par leurs distributions de couleur (modalité #2)

La figure 4.15 montre quelques exemples d'images acquises pour cette modalité. Les difficultés proviennent de l'encombrement de la scène dû éventuellement à la présence de plusieurs individus, des changements d'illumination suivant les lieux traversés, et les changements d'apparence, des pertes temporaires ou des sauts de dynamique de la cible dans le flot vidéo. La distance relative H/R ($< 3m$) permet de s'appuyer ici encore sur la silhouette de la cible dans la fonction de mesure. Toute fonction de mesure reposant sur l'attribut mouvement est écartée pour cette modalité où le robot est mobile. Nous privilégions ici la fonction de mesure $F2C1$, fusionnant forme (contour de tête) et distributions de couleur, car les évaluations du § 3.5 ont montré que cette mesure offre, à cette distance, un bon compromis en termes de précision, pouvoir discriminant et temps de traitement. Une extension logique est de considérer plusieurs régions d'intérêt distinctes spatialement et colorimétriquement. Les distributions de ces sous-régions sont prises en compte par la relation (3.9).

Typiquement, nous ajoutons une seconde distribution de couleur liée aux vêtements afin de permettre la différenciation du sujet guidé lorsque plusieurs individus sont dans le champ de vue. La gestion de plusieurs sous-régions limitent en outre les éventuelles dérives lors du suivi, notamment lorsqu'on met en place une mise à jour des distributions de référence (équation (3.10)) pour coller à l'apparence courante de la cible. Ces dérives sont d'autant mieux contrôlées que l'attribut forme est fusionné dans la fonction de mesure ce qui permet de recaler le *template* sur les contours de la forme avant d'effectuer la mise à jour de la distribution de couleur.

Pour cette modalité, cette mise à jour est ici justifiée car les mouvements réels de la cible et les conditions d'illumination rencontrées dans les lieux traversés peuvent induire des changements d'apparence conséquents. Le coefficient α pondère l'influence dans la mise à jour de la distribution $h^c_{E[x_k]}$ liée à l'état courant estimé de la cible et de la distribution de référence $h^c_{ref,k-1}$. Ce coefficient est déterminé empiriquement, et fixé à 0.3 (resp. 0.1) pour la distribution relative au visage (resp. au torse).

Une alternative plus naturelle en vue de discriminer la personne guidée est d'isoler son visage *via* un processus de détection et de classification de visages. Cette information est alors prise en compte dans la fonction d'importance. Nous avons repris le classifieur initialement développé par l'Institut des Systèmes et de Robotique (ISR) de Coimbra.[5]

[5] dans le cadre de la coopération franco-portugaise GRICES portant sur le thème de l'interaction visuelle et financée depuis 2003 par la DRI du CNRS.

Ce classifieur, intégré sur la plateforme Rackham, n'est pas à ce jour pris en compte dans nos fonctions de suivi car il doit encore être évalué seul.

L'apprentissage repose sur une analyse en composantes principales d'un ensemble d'imagettes, de résolution $N = W * H$, extraites hors-ligne par le détecteur de visages pour un sujet donné. Les M vecteurs propres Φ_i, $(i = 1, \ldots, M)$ associés aux M valeurs propres les plus élevées définissent alors un sous-espace vectoriel $F = \{\Phi_i\}_{i=1}^{M}$. Soit $\bar{F} = \{\Phi_i\}_{i=M+1}^{N}$ son sous-espace orthogonal. La reconnaissance est ensuite effectuée en ligne. La règle de décision et la confiance associée à celle-ci reposent alors sur la projection d'une image issue du détecteur sur ces deux sous-espaces. Le lecteur peut se référer à [Menezes et al., 2004] pour plus de détails. La figure 4.14 montre quelques exemples de détections et de reconnaissances dans le flot vidéo.

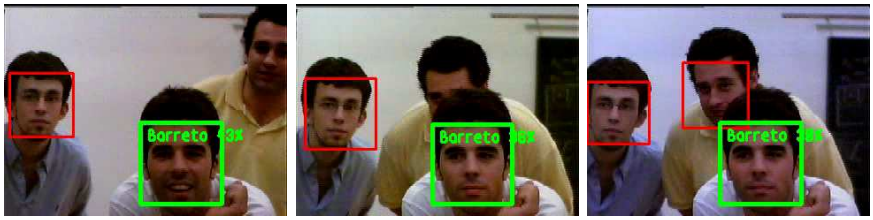


FIG. 4.14 – Exemple de : détections de visages (en rouge), de détections/reconnaissances (en vert)

4.3.3 Fonctions d'importance et stratégies de filtrage envisagées

Les stratégies basées sur des fonctions d'importance reposant sur la dynamique seule semblent *a priori* inadaptées pour cette modalité car il faut gérer les décrochages éventuels de la cible. Elles seront néanmoins considérées, de sorte que nous évaluerons les six stratégies de filtrage proposées précédemment. Pour les stratégies FIM et $FIDM$, la fonction d'importance exploitera, par un mélange de Gaussiennes, les résultats des deux détecteurs de visages et de *blobs* « peau ». Cette association notée $FSBD$ au § 3.5 minimise les faux négatifs (voir figure 3.37).

Pour ces fonctions d'importance et de mesure, les six stratégies de filtrage déclinées dans le tableau 4.1 sont évaluées sur des séquences cohérentes avec cette modalité.

4.3.4 Évaluation des stratégies de suivi envisagées

Nous suivons le protocole d'évaluation décrit au § 4.1.5. Les évaluations portent sur 20 séquences (20×20 réalisations au total) plus ou moins complexes et représentatives de cette deuxième modalité. La figure 4.15 montre quelques exemples d'images tirées de ces séquences.

Certains événements - *e.g.* des sauts dans la dynamique, des fausses détections ou des non-détections de la cible - peuvent engendrer un décrochage du filtre. Considérons tout d'abord le sous-ensemble des séquences-types qui ne donnent pas lieu à ces décrochages. Pour ces séquences « ordinaires », les stratégies ont alors un comportement très similaire.

FIG. 4.15 – Exemples d'images acquises en interaction proche (modalité #2)

La figure 4.16 montre les précisions et taux d'échec obtenus, respectivement de l'ordre de 10 pixels et de 1% à partir de 100 particules.

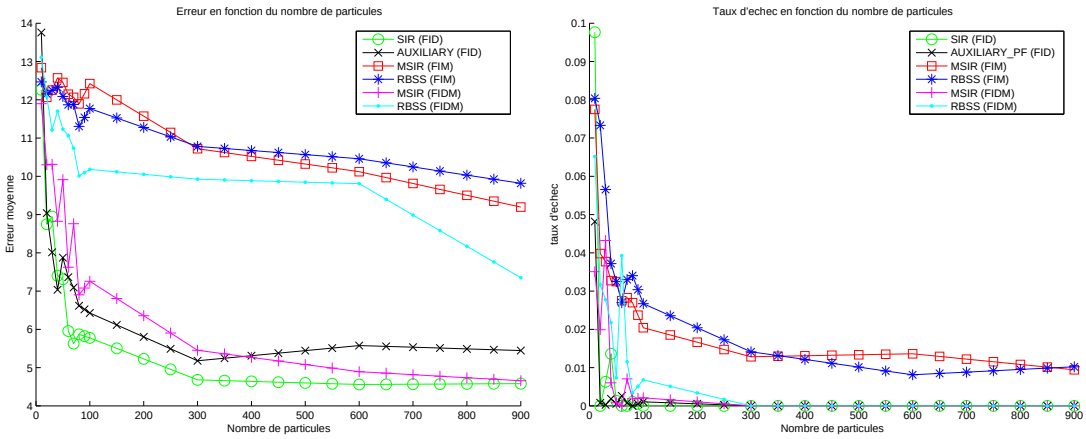


FIG. 4.16 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences « ordinaires » pour les stratégies envisagées (modalité #2)

La figure 4.17 montre une réalisation sur une séquence-type pour la stratégie $FIDM1$. Pour chaque image, la particule rouge est la moyenne *a posteriori* estimée par le filtre.

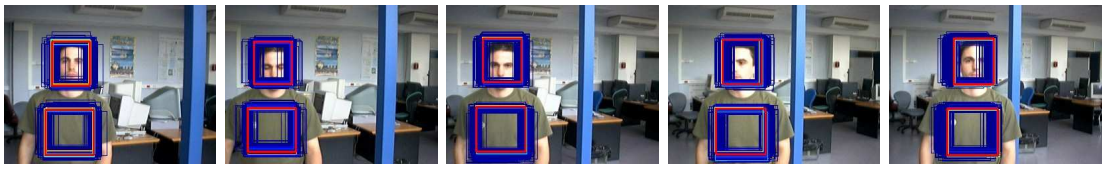


FIG. 4.17 – Exemple de réalisation sur une séquence « ordinaire » pour la stratégie $FIDM1$ (modalité #2)

En l'absence de sauts de dynamique et de fausses détections, il est montré que les

différentes stratégies restent robustes aux changements d'illumination. La mise à jour de la distribution de couleur permet de s'adapter aux changements d'apparence de la cible et de garantir le suivi grâce à la fusion des deux attributs forme et couleur dans la fonction de mesure. La figure 4.18 montre, pour la stratégie $FIDM2$, un exemple de suivi sur une séquence type où l'éclairage varie de façon significative, ici dans un couloir.

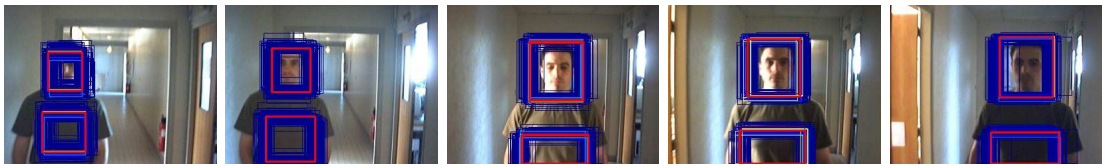


FIG. 4.18 – Exemple de réalisation sur une séquence incluant des variations d'illumination pour la stratégie $FIDM2$ (modalité #2)

Des conditions d'illumination plus extrêmes peuvent aboutir à la non-détection de la cible. Les comportements de nos six filtres restent néanmoins similaires et proches d'une stratégie FID. En effet, les stratégies $FIDM$ s'appuient alors sur la seule dynamique dans leurs fonctions d'importance tandis que pour les stratégies FIM, l'estimation est réalisée par CONDENSATION classique. En l'absence de sauts de dynamique ou de fausses détections, la modalité #2 ne justifie donc pas l'utilisation de stratégies particulières, la robustesse est assurée par la seule fonction de mesure qui est suffisamment discriminante et précise (§ 3.5). Dans ce cadre, la figure 4.19 montre les précisions et taux d'échec obtenus pour les séquences incluant de forts changements d'apparence.

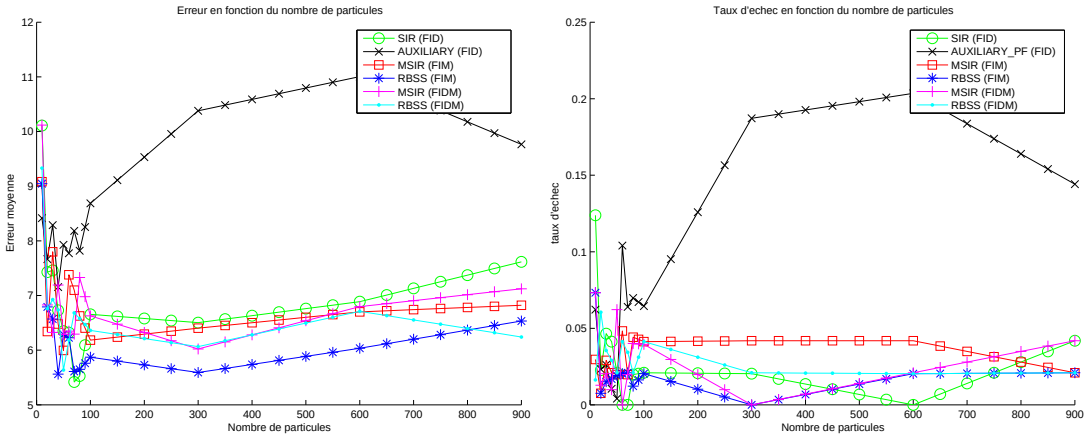


FIG. 4.19 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences incluant des changements d'apparence pour les stratégies envisagées (modalité #2)

La figure 4.20 illustre le suivi pour une de ces séquences mais incluant de forts changements d'apparence pour une stratégie $FIM1$.

En présence (inéluctable) de sauts dans la dynamique de la cible, les évaluations

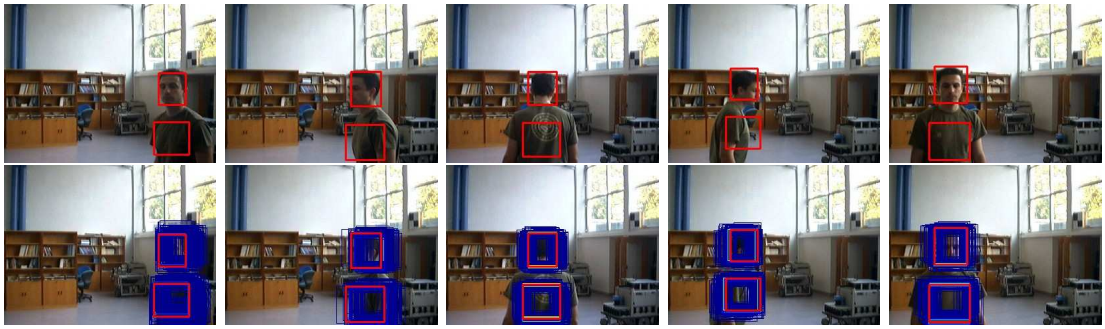


FIG. 4.20 – Exemple de réalisation sur une séquence incluant des changements d'apparence pour une stratégie $FIM1$ avec tracé de la particule moyenne *a posteriori* - haut -, de toutes les particules - bas - (modalité #2)

montrent logiquement la supériorité des stratégies FIM et $FIDM$ sur les stratégies FID où des pertes de cibles sont alors observées. La figure 4.21 montre des réalisations pour ces différentes stratégies sur des séquences incluant des sauts dans la dynamique. Le modèle de dynamique seul ne permet pas de repositionner correctement les particules autour de la cible pour une stratégie FID. Pour les stratégies FIM ou $FIDM$, ce repositionnement est rendu possible grâce à la détection.

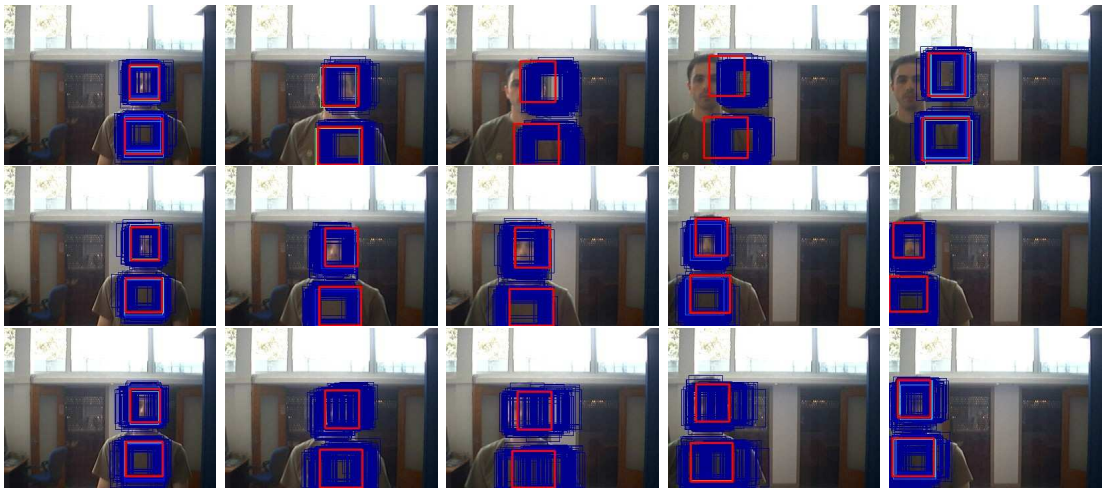


FIG. 4.21 – Exemple de réalisation pour une séquence incluant des sauts dans la dynamique de la cible pour les stratégies $FID1$ - haut -, $FIM1$ - milieu - et $FIDM1$ - bas - (modalité #2)

Une évaluation systématique sur les séquences de ce type est effectuée. La figure 4.22 compare, pour nos stratégies, les erreurs et taux d'échec obtenus.

Les taux d'échecs obtenus (15 à 20%) sur plusieurs réalisations pour les stratégies FID sont largement supérieurs à ceux obtenus pour les stratégies FIM et $FIDM$ ($\simeq 2\%$). Les erreurs moyennes sont logiquement dégradées pour les stratégies FID.

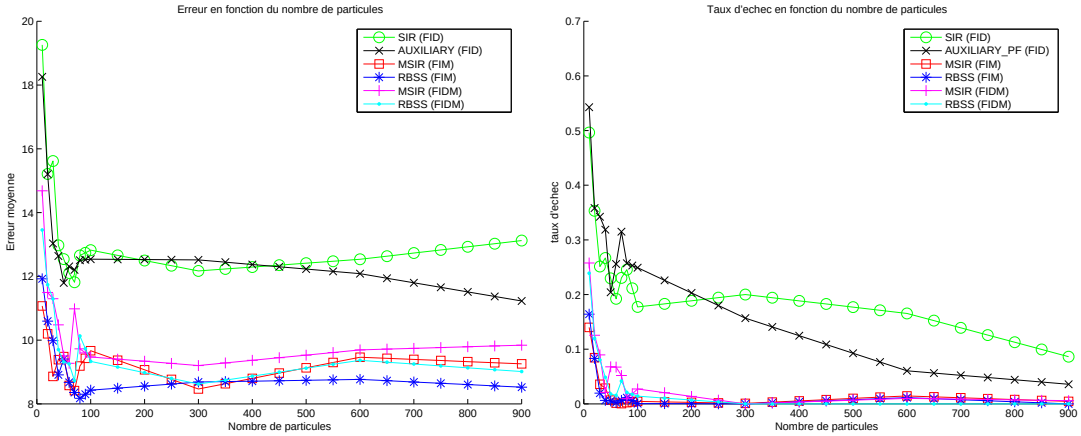


FIG. 4.22 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences incluant des sauts dans la dynamique pour les stratégies envisagées (modalité #2)

Dans cette modalité #2, la personne guidée peut à cette distance, sortir du champ de vue de la caméra. Les pertes de cibles sont à prendre à considération dans nos stratégies de filtrage. Les stratégies FIM et $FIDM$, par leur capacité de ré-initialisation, semblent adaptées. L'évaluation des six stratégies sur plusieurs séquences de ce type donne des précisions et taux d'échec similaires à ceux illustrés sur la figure 4.22. Néanmoins, le taux d'échec des stratégies FID est légèrement inférieur que précédemment car la cible sortie du champ de vue réapparaît le plus souvent au même endroit, permettant ainsi au suivi de « raccrocher » la cible.

La perspective de voir notre robot guider les visiteurs dans un musée nous amène à considérer plusieurs personnes dans le champ de vue. Pour cette modalité de guidage, la fonction de suivi ne doit pas se laisser distraire par les autres usagers de l'environnement. Ceux-ci peuvent tout aussi bien apparaître ponctuellement dans le champ de vue ou, ce qui est alors plus problématique, occulter ponctuellement la cible. L'identité de la cible est déclinée par la seconde distribution de couleur liée aux vêtements (figure 4.13). Une variante, plus intuitive et moins restrictive, porte sur l'intégration à venir d'un classifieur de visages (§ 4.3.2) dans nos filtres. La classification probabiliste de chaque visage détecté sera alors prise en compte dans les fonctions d'importance des filtres.

En présence de plusieurs individus, les stratégies FID permettent de dissocier la cible si les dynamiques associées à leurs mouvements apparents et/ou la seconde distribution de couleur sont incompatibles avec celles de la cible. Concernant les stratégies FIM et $FIDM$, leurs fonctions d'importance positionnent *a priori* les particules sur tous les individus détectés puis la différenciation s'effectue avec les mêmes critères que précédemment. Les stratégies FIM sont logiquement en échec lorsque la cible n'est pas détectée alors que les autres individus le sont. Cette situation très particulière est rencontrée lorsque la personne suivie est de dos, alors que le regard d'au moins un autre individu est dirigé vers la caméra, donc donnant lieu à détection d'un visage ou de régions peau. La figure 4.23 illustre ce cas de figure pour les stratégies $FIDM1$ et

$FIM1$.Cette dernière est clairement mise en échec.

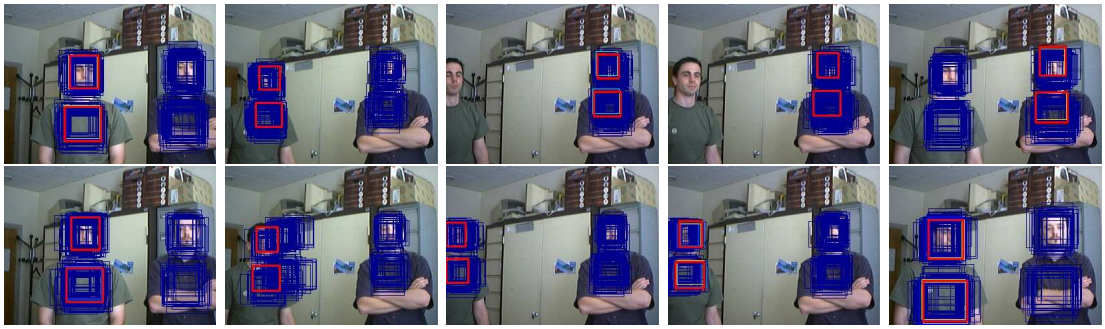


FIG. 4.23 – Exemple de réalisation sur une séquence incluant deux personnes pour les stratégies $FIM1$ - haut - et $FIDM1$ - bas -. La cible est l'individu situé à gauche (modalité #2)

La figure 4.24 compare, pour les différentes stratégies, les erreurs moyennes et taux d'échec obtenus sur des séquences contenant deux personnes.

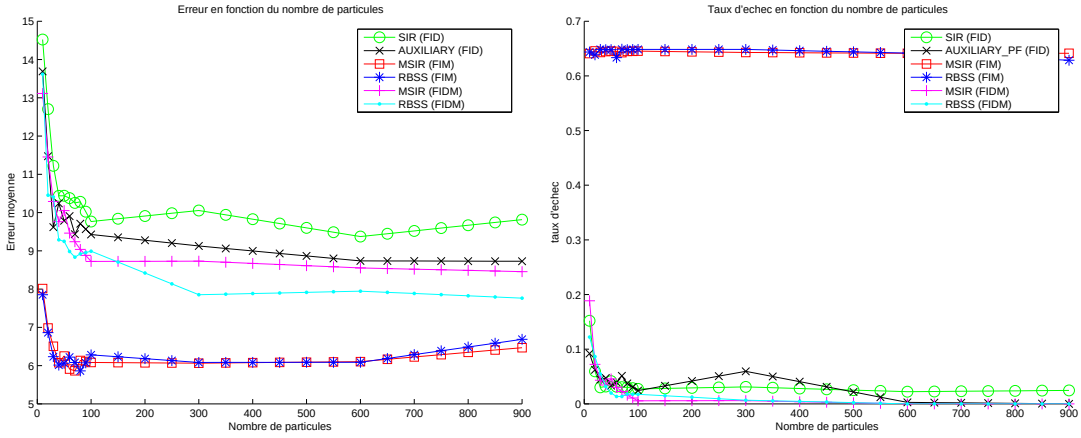


FIG. 4.24 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences incluant deux individus pour les stratégies envisagées (modalité #2)

Forts des considérations précédentes, les stratégies FIM donnent des taux d'échec plus élevés ($\simeq 65\%$) que les stratégies FID et $FIDM$. En revanche, dans le cas de non-décrochages du filtre, elles donnent de meilleures précisions, devant les stratégies $FIDM$ puis les stratégies FID.

Un dernier cas de figure concerne l'occultation de la cible dans le flot vidéo, notamment par un individu autre qui passe momentanément entre le robot et le visiteur guidé. Les stratégies FID conduisent à des échecs, le filtre commutant alors sur l'individu situé au premier plan malgré le modèle de vêtement. Pour les autres stratégies, les particules placées selon la détection assurent la continuité dans le suivi. La figure 4.25 montre un exemple de réalisation dans ce scénario pour les stratégies FID et $FIDM$.

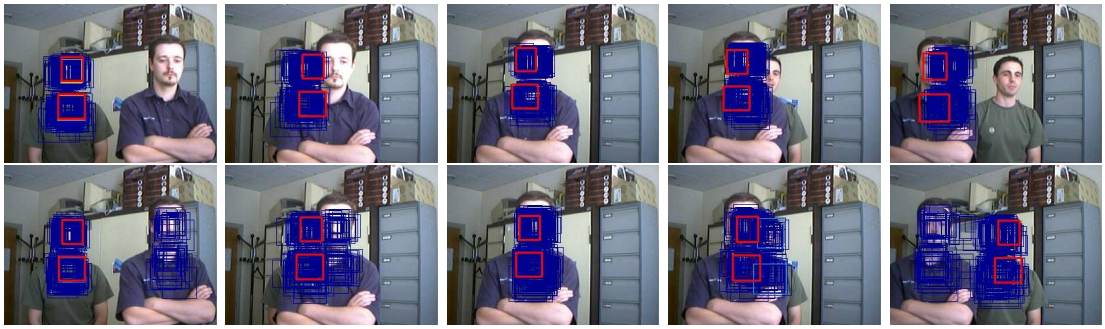


FIG. 4.25 – Exemple de réalisation en présence de deux personnes avec FID - haut - et $FIDM$ - bas -. La cible est la personne située à gauche dans les images (modalité #2)

Des évaluations plus globales sur les séquences exhibant ces situations sont reportées sur la figure 4.26. Les stratégies FID donnent un taux d'échec nettement supérieur à celui des autres stratégies. On notera tout de même que la stratégie $FID2$ voit son taux d'échec décroître significativement avec le nombre de particules. La précision reste, quant à elle, comparable quelle que soit la stratégie envisagée.

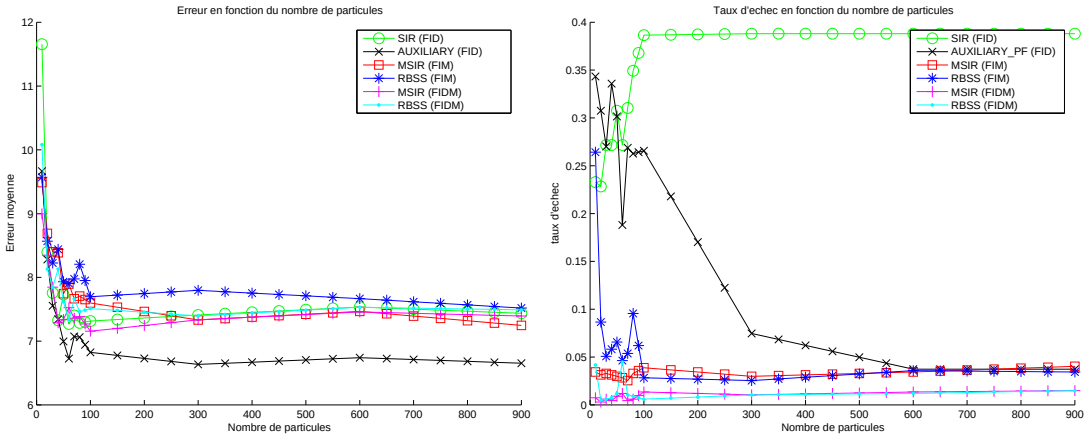


FIG. 4.26 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences incluant des occultations de cibles pour les stratégies envisagées

4.3.5 Discussion

Nous avons proposé et évalué différents filtres dédiés au suivi à partir d'une caméra embarquée sur un robot mobile censé guider un individu donné dans un environnement d'intérieur. La mise en œuvre sur notre robot de cette modalité d'interaction suppose que la stratégie de filtrage retenue soit robuste aux diverses situations et artefacts rencontrés durant l'éxécution de cette modalité. L'apparence de la cible peut changer de par les conditions d'illumination propres aux lieux visités ou aux mouvements réels de l'individu

guidé. La dynamique inter-images associée est difficilement caractérisable et peut inclure des sauts. Ces sauts, de même que les déplacements du robot lors du guidage, peuvent aboutir à la non-observation momentanée de la cible dans le flot image. Enfin, le robot partage l'environnement avec plusieurs usagers qui peuvent être ponctuellement présents ponctuellement dans le champ de vue ou masquer temporairement l'individu guidé. La stratégie retenue au final doit être robuste à ces situations variées et plus ou moins complexes à gérer.

Nous avons donc caractérisé le comportement de nos filtres pour les situations répertoriées ci-dessus. Notre étude est illustrée par des réalisations de suivi sur des séquences relatives à une situation donnée. Des évaluations chiffrées ont été proposées et commentées. Elles portent sur deux critères : la précision et la robustesse.

Grâce à une fonction de mesure adéquate, les six stratégies sont robustes aux changements d'apparence de la cible. Lors de sauts dans sa dynamique ou de sortie du champ de vue, les évaluations montrent clairement un avantage pour les stratégies FIM et $FIDM$. Néanmoins, les évaluations mettent en évidence qu'une fonction d'importance exclusivement basée sur la mesure (FIM) pose problème en cas de non détection de la cible suivie. La figure 4.27 compare les temps de traitement obtenus sur l'ensemble des séquences pour les différentes stratégies.

Nous optons finalement pour les stratégie $FIDM$, dont les comportements sont les plus satisfaisants pour l'ensemble des situations évaluées. Les particules placées selon la détection permettent une réinitialisation en cas de perte de la cible et limitent les risques de décrochage dûs à des sauts dans sa dynamique. Quand aux particules distribuées suivant la dynamique, elles assurent un suivi de la cible lorsque celle-ci n'est pas détectée.

Parmi les stratégies $FIDM$, l'algorithme de filtrage RBSS est légèrement plus performant que l'algorithme MSIR en termes de précision et taux d'échec. Cet avantage peut s'expliquer par le rééchantillonage intermédiaire (chapitre 2), qui permet une association plus cohérente, vis à vis de la dynamique, entre les particules placées selon la mesure à l'instant k et leurs particules « parents » à l'instant $k-1$. On notera enfin sur la figure 4.27 que les temps de traitement obtenus pour les deux stratégies $FIDM$ sont très similaires pour le nombre de particules couramment utilisé (entre 100 et 200).

Nous privilégions donc pour cette modalité d'interaction, la stratégie $FIDM2$.

4.4 Suivi pour la surveillance

4.4.1 Considérations générales

Les distances relatives H/R ($> 3m$) dans cette modalité ne permettent pas d'exploiter la silhouette du *template* tel que nous l'avons défini. La cible est naturellement modélisée ici par son rectangle englobant dont l'état x_k est caractérisé par les seules position (u_k, v_k) et échelle s_k. Nous utilisons ici encore un modèle de dynamique de type marche aléatoire. A cette distance, les mouvements apparents de l'individu suivi sont moins rapides. Nous fixons empiriquement les bruits de dynamique comme suit :

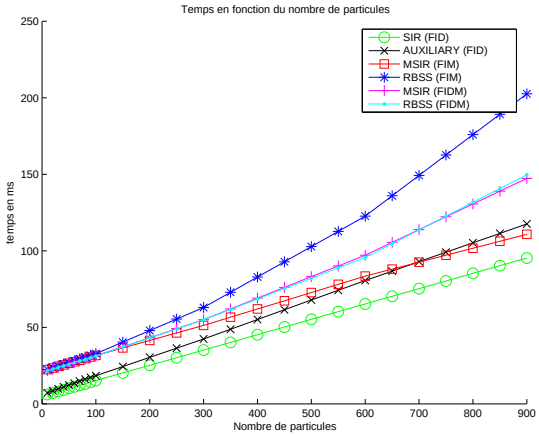


FIG. 4.27 – Temps moyens de traitement *vs* nombre de particules sur l'ensemble des séquences pour les différentes stratégies (modalité #2)

$(\sigma_u, \sigma_v, \sigma_s) = (6, 3, 2.10^{-2})$. Nous notons pour la suite :

$$x_k = [u_k, v_k, s_k]'.$$

D'autres modèles de dynamique ont été envisagés par le passé, citons un modèle AR du $2^{ème}$ ordre pour la modélisation de mouvements apparents à vitesse constante, et sont peut-être tout aussi réalistes pour cette modalité. Même si nos évaluations se limitent ici encore aux seules stratégies de filtrage, nous envisageons de les étendre aux modèles de dynamique.

Pour définir plus finement ces modèles, nous pourrons exploiter la carte connue *a priori* de l'environnement. Dans cette modalité dite de surveillance, le robot pourra se positionner à des endroits stratégiques du site afin d'observer « au mieux » les lieux de passage privilégiés par les usagers, par exemple les chemins entre deux portes répertoriées dans la carte. Ces considérations devraient nous aider à choisir un modèle de dynamique éventuellement plus adapté à cette modalité.

4.4.2 Fonctions de mesure envisagées

La figure 4.29 montre quelques images acquises pour cette modalité. Les difficultés proviennent de l'encombrement de la scène, dû éventuellement à plusieurs individus, ainsi que des pertes ou des mouvements apparents saccadés de la cible dans le flot vidéo.

Nous exploitons alors l'apparence de la cible dans une fonction de mesure basée sur le calcul de la distribution de couleur contenue dans le rectangle englobant la personne. Celle-ci, notée $C1$ dans les évaluations du § 3.5, est assez discriminante et peu coûteuse en temps de calcul, car les distributions sont calculées sur des régions image de petites tailles. Enfin, sa relative imprécision est compatible avec cette modalité d'interaction et les distances relatives H/R considérées.

Le robot étant à l'arrêt, nous fusionnons ici cette mesure avec une mesure de distribution du mouvement dans une région centrée sur la cible. La figure **??** montre ces deux régions d'intérêt, la plus petite étant relative à la distribution de couleur et la plus grande ayant trait à la distribution de mouvement. On englobe ainsi au mieux les sous-régions mobiles qui se situent majoritairement à la périphérie de la cible.

4.4.3 Fonctions d'importance et stratégies de filtrage envisagées

Les fonctions d'importance envisagées auparavant ne peuvent pas être exploitées ici vu les distances relatives H/R considérées. La fonction d'importance reposant sur la détection de mouvement (notée MD au § 3.2.2) demeure la seule alternative dans ce contexte de surveillance. Concernant les stratégies de filtrage, le risque élevé d'occultations nous oriente naturellement vers les stratégies $FIDM$, jugées plus performantes dans les évaluations précédentes. Nous nous focaliserons donc sur l'évaluation des stratégies $FIDM1$ et $FIDM2$ pour cette modalité. Pérez *et al.* dans [Pérez et al., 2004] exploitent les mêmes mesures et une stratégie de filtrage hiérarchique notée $HIERARC$ dans un contexte de surveillance (§D-3). Cette stratégie qui intègre successivement les informations de mouvement puis de couleur sera donc logiquement évaluée ici et comparée aux deux stratégies $FIDM$.

4.4.4 Évaluation des stratégies de filtrage envisagées

Nous suivons le protocole d'évaluation décrit au § 4.1.5. Les 20 séquences traitées (donc 20×20 réalisations au total), plus ou moins complexes, sont représentatives de cette dernière modalité. La figure 4.29 montre quelques images tirées de ces séquences.

FIG. 4.29 – Exemples d'images acquises en surveillance (modalité #3)

Considérons tout d'abord le sous-ensemble des séquences « ordinaires » *i.e.* sans occultation ou arrêt de la cible. La figure 4.30 montre une réalisation sur une séquence-type pour les stratégies $FIDM1$ et $HIERARC$.

Les filtres s'initialisent correctement sur l'individu entrant dans le champ de vue, et son suivi s'effectue avec succès. Nous constatons, pour la stratégie $HIERARC$ que le nuage de particules (en bleu sur la figure) est plus concentré sur la cible. Cette bonne distribution des particules se répercute naturellement sur la précision du filtre. La figure 4.31 montre les erreurs et taux d'échec obtenus sur un ensemble de séquences

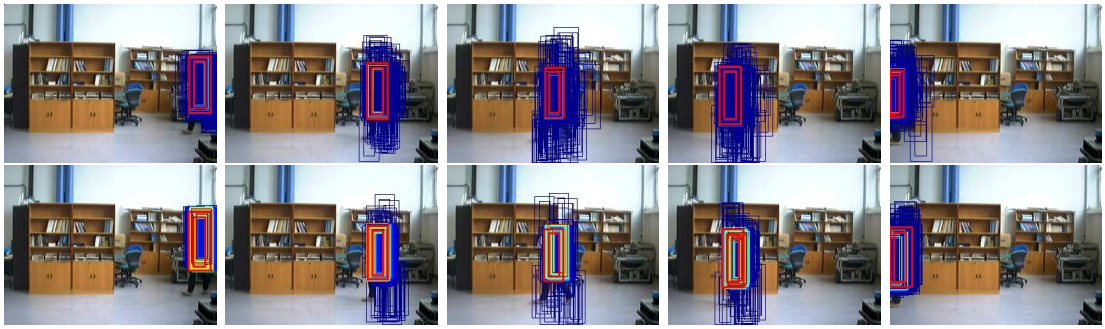


FIG. 4.30 – Exemples de réalisations sur une séquence « ordinaire » avec $FIDM1$ en haut et $HIERARC$ en bas (modalité #3).

similaires pour les trois stratégies. Ces évaluations montrent que les erreurs sont plus faibles pour la stratégie $HIERARC$ tandis que les taux d'échec sont extrêmement faibles pour les trois stratégies.

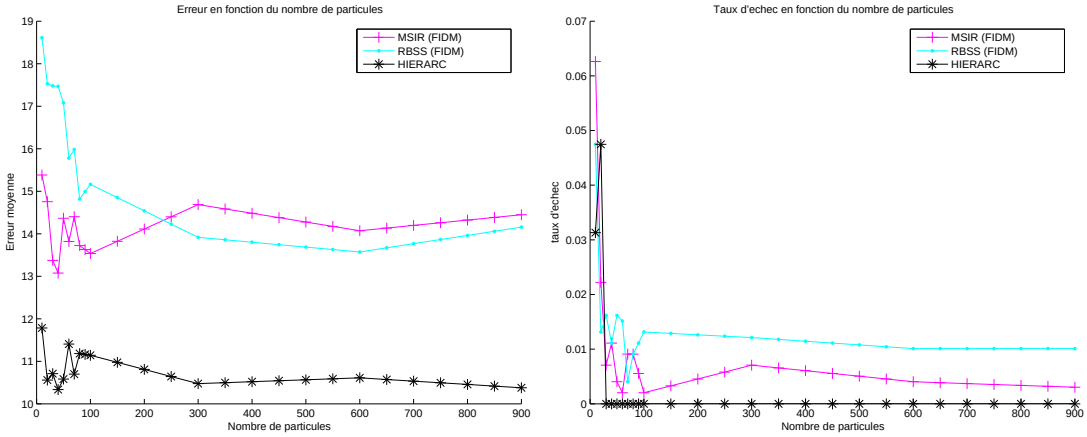


FIG. 4.31 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences de suivi « ordinaire » pour les stratégies $FIDM1$ et $HIERARC$ (modalité #3)

Analysons les situations difficiles qui peuvent engendrer des décrochages de nos filtres. En premier lieu, l'individu suivi peut être en mouvement intermittent. La figure 4.32 montre le comportement du filtre $FIDM1$ sur une séquence incluant un arrêt de la cible. Le pourcentage de particules placées selon la dynamique par la fonction d'importance, et la fonction de mesure, suffisamment discriminante, permettent de suivre la cible avec succès.

La figure 4.33 compare, pour les trois stratégies, les erreurs moyennes et taux d'échecs relatifs aux séquences incluant un arrêt de la cible. Les résultats obtenus confirment l'avantage de la stratégie $HIERARC$ pour la précision tandis que les taux d'échec sont plus faibles que précédemment et ce quelle que soit la stratégie.

La distance d'interaction (>3m) nous conduit à prendre en considération d'éven-

FIG. 4.32 – Exemple de réalisation sur une séquence incluant un arrêt de la cible pour la stratégie $FIDM1$ (modalité #3)

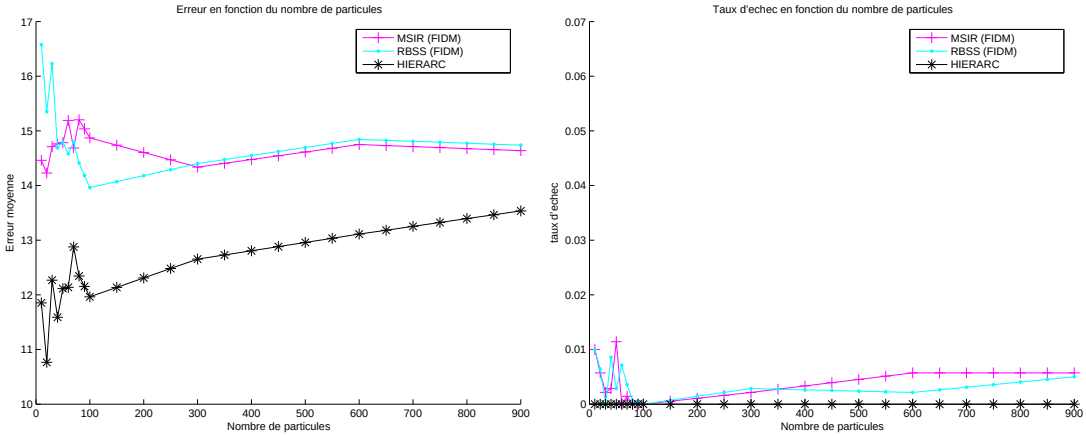


FIG. 4.33 – Erreurs et taux d'échec *vs.* nombre de particules sur des séquences incluant un arrêt de la cible pour les stratégies $FIDM$ et $HIERARC$ (modalité #3)

tuelles occultations de la cible. Elles sont engendrées par des objets ou d'autres individus présents dans la scène. La figure 4.34 illustre le comportement des stratégies $FIDM1$ et $HIERARC$ sur une séquence incluant une occultation durable de la cible par un objet. La stratégie $HIERARC$ est, dans ce scénario, mise en échec tandis que la stratégie $FIDM1$ permet de « raccrocher » la cible après occultation. Pour cette dernière, certaines particules sont redistribuées selon la détection ($q(x_0)$) (donc sur la cible détectée lors de sa réapparition) et ceci indépendamment de la dynamique. Pour la stratégie $HIERARC$, les particules positionnées selon la détection sont éliminées lors du premier rééchantillonnage car trop incohérentes du point de vue de la dynamique qui privilégie l'état avant occultation.

L'évaluation sur plusieurs séquences de ce type confirment ces observations. La figure 4.35 compare les erreurs moyennes et taux d'échecs obtenus. Les taux d'échec obtenus sont clairement plus élevés pour la stratégie $HIERARC$.

Considérons maintenant des occultations engendrées par d'autres individus. La figure 4.37 montre une réalisation sur une séquence incluant le croisement de deux individus pour la stratégie $HIERARC$. Une stratégie $FIDM$ permet également le suivi sans décrochage de la cible. Par les fonctions d'importance de ces schémas de filtrage, les particules sont placées initialement sur les deux cibles potentielles mais leurs dynamiques respectives distinctes et la mesure finale pour la mise à jour des poids permettent de

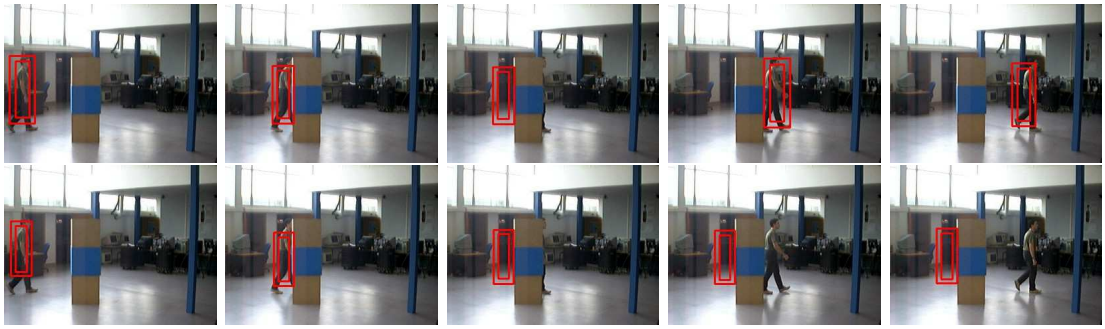


FIG. 4.34 – Exemple de réalisation sur une séquence incluant une occultation longue et totale de la cible pour les stratégies $FIDM1$ en haut et $HIERARC$ en bas (modalité #3)

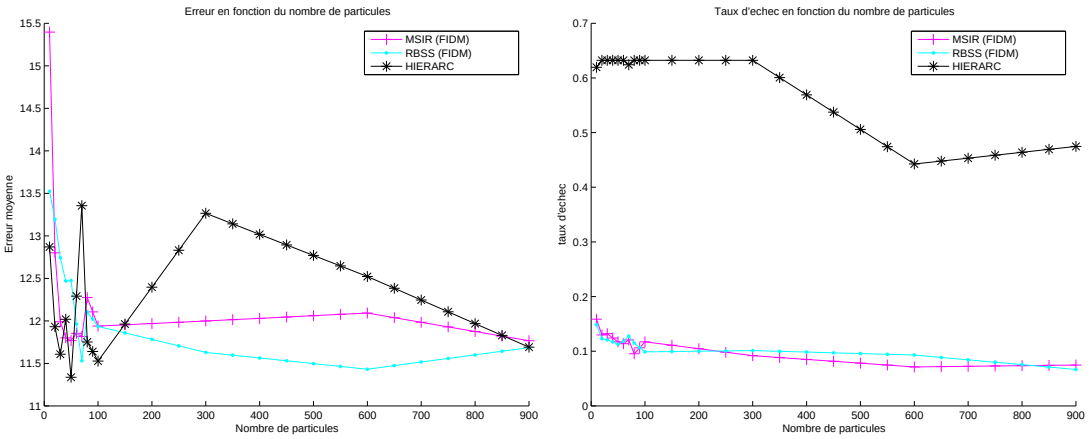


FIG. 4.35 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences incluant des occultations importantes de la cible pour les stratégies $FIDM1$ et $HIERARC$ (modalité #3)

raccrocher la cible après occultation.

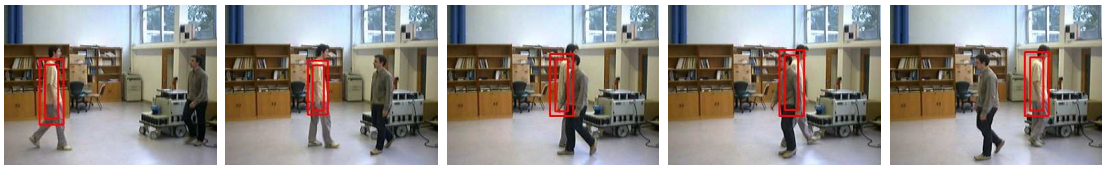


FIG. 4.36 – Exemple de réalisation sur une séquence incluant le croisement de deux personnes pour la stratégie $HIERARC$ (modalité #3)

L'évaluation sur plusieurs séquences de ce type montre que les trois stratégies ont des performances similaires et satisfaisantes : les erreurs moyennes et les taux d'échec obtenus sont négligeables (figure 4.37). Ces faibles taux s'expliquent par la nature brève

de l'occultation qui ne génère pas d'incohérence dans la dynamique de la cible et ne contribue pas au modèle de couleur pendant suffisament de temps pour entraînner sa dérive.

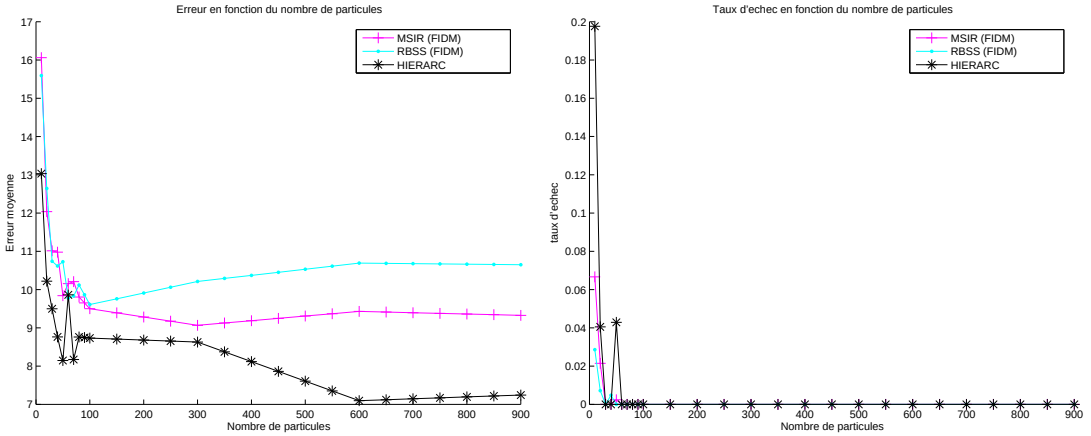


FIG. 4.37 – Erreurs et taux d'échec *vs* nombre de particules sur des séquences incluant un croisement d'individus pour les stratégies $FIDM1$ et $HIERARC$ (modalité #3)

Une variante, dans ce scénario[6], est de supposer la cible immobile. Ce scénario est illustré par la figure 4.38 pour les stratégies $FIDM1$ et $HIERARC$. La stratégie $HIERARC$ est logiquement en échec. En effet, le premier rééchantillonnage va privilégier les particules associées à des zones, certes mobiles, mais cohérentes en terme de dynamique. En début de suivi, la cible est isolée, la dynamique permet de rester focalisé sur la cible. Lorsque le second indivivu vient occulter la cible, non seulement il est vraisemblable du point de vue de la mesure basée sur le mouvement mais sa dynamique devient compatible avec la dynamique *a priori* de la cible et par conséquent, le filtre s'accroche sur lui. Une stratégie $FIDM$ permet, quant à elle, la réinitialisation correcte du suivi après croisement.

L'évaluation sur plusieurs séquences relatives à ce scénario est illustrée par la figure 4.39 qui compare les erreurs moyennes et taux d'échec obtenus pour les trois stratégies. Les taux d'échec sont nettement supérieurs pour la stratégie $HIERARC$ et confirment les conclusions effectuées sur l'exemple 4.38.

Les situations précédentes peuvent être simultanément rencontrées si l'on considère plusieurs individus dans le champs de vue. La figure 4.40 montre une réalisation sur une séquence incluant un groupe de personnes pour la stratégie $FIDM1$.

La figure 4.41 compare les erreurs et taux d'échecs pour un suivi incluant un groupe de personnes. Les erreurs moyennes sont plus faibles pour la stratégie $HIERARC$. Les taux d'échec sont par contre meilleurs avec les stratégies $FIDM$.

[6] *a priori* courant pour notre robot-guide.

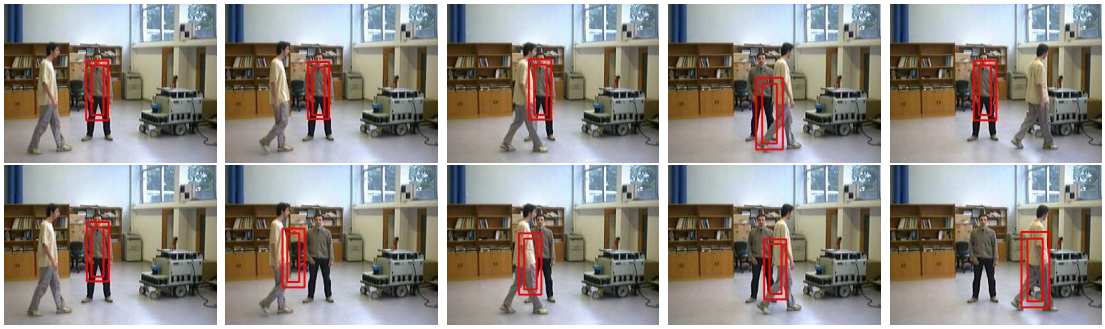


FIG. 4.38 – Exemple de réalisation sur une séquence incluant une occultation de la cible supposée statique par un autre individu pour les stratégie $FIDM1$ en haut et $HIERARC$ en bas (modalité #3)

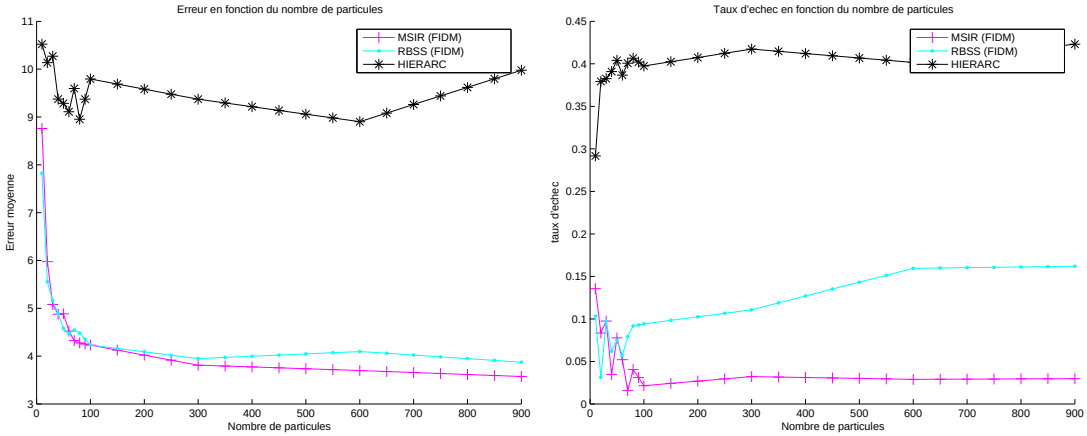


FIG. 4.39 – Erreurs et taux d'échec *vs.* nombre de particules sur des séquences incluant une cible statique occultée par un autre individu pour les stratégies $FIDM$ et $HIERARC$ (modalité #3)

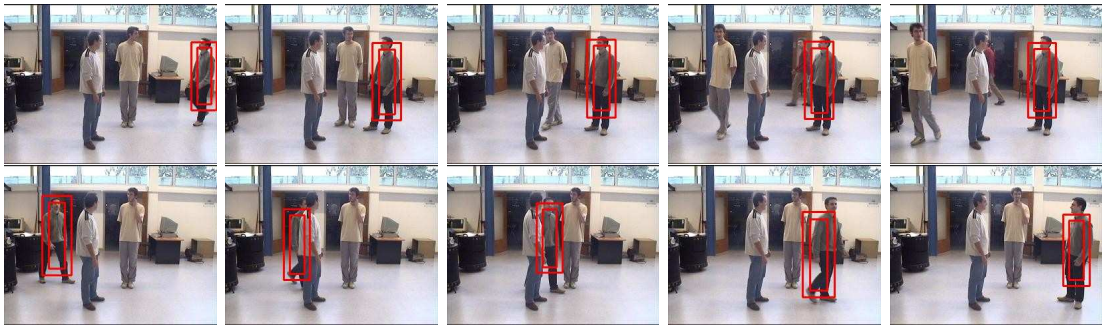


FIG. 4.40 – Exemple de réalisation sur une séquence incluant un groupe de personnes pour la stratégie $FIDM1$ (modalité #3)

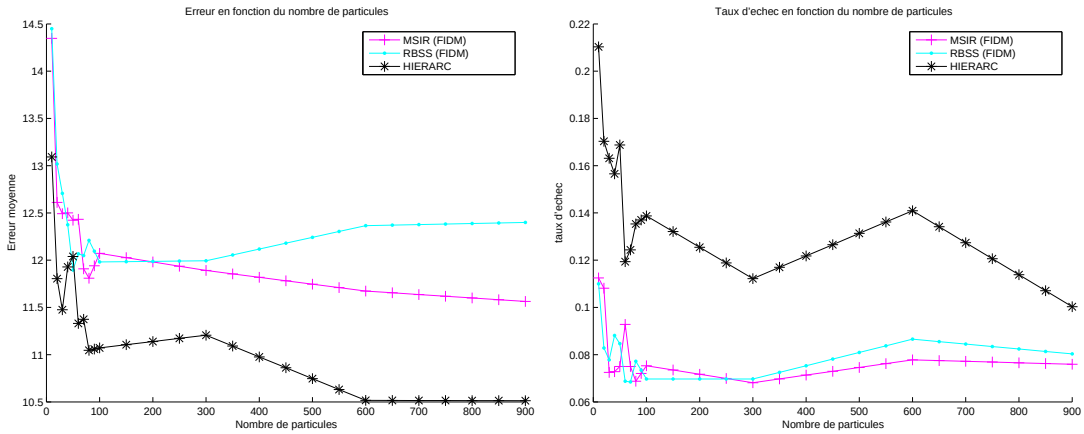


FIG. 4.41 – Erreurs et taux d'échec *vs.* nombre de particules sur des séquences incluant un groupe de personnes pour les stratégies $FIDM$ et $HIERARC$ (modalité #3)

4.4.5 Discussion

Nous avons évalué les stratégies $FIDM$ et $HIERARC$ pour cette troisième modalité. Le but pour le robot, ici à l'arrêt, consiste à « surveiller » les usagers présents dans l'environnement et donc susceptibles d'interagir avec lui. La stratégie de filtrage implantée sur le robot doit être robuste dans des scenarii variés et plus ou moins complexes rencontrés dans ce contexte.

Pour les scenarii répertoriés, nous avons donc caractérisé le comportement des stratégies $FIDM1, FIDM2$ et $HIERARC$. L'algorithme de filtrage hiérarchique $HIERARC$ conduit globalement à des précisions plus satisfaisantes, mais les stratégies $FIDM$ aboutissent à des taux d'échec très inférieurs dans plusieurs situations typiques de ce contexte. Nous privilégierons donc les stratégies $FIDM$ pour leur robustesse. Ce critère nous semble essentiel car l'enjeu est, dans cette modalité, de suivre sans décrochage un individu parmi N.

Parmi les stratégies $FIDM$, les algorithmes de filtrage RBSS et MSIR ont globalement des performances très semblables et nous amènent à choisir indifféremment l'un ou l'autre.

Les temps de traitement obtenus pour ces deux stratégies sont également très similaires et ne permettent pas de les différencier (figure 4.42). On notera néanmoins, que l'algorithme hiérarchique est un peu moins coûteux en temps de traitement pour le nombre de particules couramment utilisé (entre 100 et 200).

4.5 Conclusion

Nous avons présenté notre plateforme Rackham dédiée à l'interaction H/R. Le défi final est de voir Rackham évoluer dans un musée, donc en présence de nombreux visiteurs. Dans ce contexte, nous avons proposé un scénario-type et décliné trois modalités

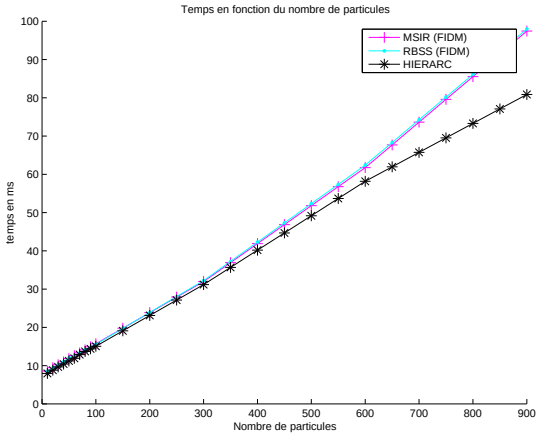


FIG. 4.42 – Temps moyens de traitement *vs* nombre de particules sur l'ensemble des séquences pour les stratégies $FIDM$ et $HIERARC$ (modalité #3)

d'interaction associées entre le robot et les visiteurs du site. Trois fonctions de suivi visuel sont nécessaires à la mise en œuvre de ces modalités sur notre plateforme. La première est relative au suivi proximal d'une personne lors de son interaction avec le robot *via* ses périphériques. Le robot est supposé à l'arrêt ici. La fonction visuelle pour la modalité #2 porte sur le suivi à mi-distance afin de guider le visiteur vers un lieu prédéfini. Enfin, la fonction visuelle pour la modalité #3 concerne la « surveillance » de lieux de passages du site afin d'interpeler les visiteurs.

Diverses stratégies de filtrage impliquant divers attributs visuels sont proposées puis évaluées sur des séquences-types. Les séquences-types sont représentatives des situations et artefacts rencontrés par le robot dans chacune de ces modalités : scènes encombrées, changements brusques d'apparence ou de dynamique de la cible, présence de plusieurs individus, occultations, etc.

Pour chaque modalité, le comportement qualitatif des filtres est illustré par des réalisations de suivi sur des séquences relatives à un scénario donné. Toutes les réalisations présentées sont intégralement accessibles à l'URL `www.laas.fr/~lbrethes`. Des évaluations chiffrées sur des jeux de séquences-types sont également proposées et commentées. Ces évaluations portent sur trois critères : précision, taux d'échec et temps de traitement. Les tableaux 4.2, 4.3 et 4.4 résument respectivement les mesures, les fonctions d'importances et les stratégies de filtrages envisagées tandis que le tableau 4.5 décline les stratégies de filtrage retenues au final.

Ces stratégies sont actuellement intégrées sur la plateforme Rackham pour une démonstration finale.

Signalons enfin que ce chapitre a contribué à la rédaction de 6 publications dans des conférences nationales [Brèthes et al., 2004b, Brèthes et al., 2006a] ou internationales [Menezes et al., 2003, Brèthes et al., 2004a, Brèthes et al., 2005, Brèthes et al., 2006b].

Fonction de mesure	Description
F1	Forme basée contours
F2	Forme basée image de distances
C1	Distribution de couleur
F1C1	Forme basée contours fusionnée à la distribution de couleur
F1M	Forme 1 combinée au mouvement (flot optique)
F2C1	Forme 2 fusionnée à la distribution de couleur
F2C2	Forme 2 combiné à la segmentation en régions de couleur peau
F1MC1	Forme 1 combinée au mouvement et fusionné avec la distribution de couleur

TAB. 4.2 – Tableau récapitulatif des fonctions de mesures.

Fonction d'importance	Description
FD	Détection de visage
FMD	Détection de visage combinée à la détection de mouvements
FSBD	Détection de visage combinée à la détection de blobs couleur
MD	Détection de mouvements
SBD	Détection de blobs couleur peau
SBMD	Détection de blobs couleur peau combinée à la détection de mouvements

TAB. 4.3 – Tableau récapitulatif des fonctions d'importance.

Stratégie de filtrage	Description
FID1	Fonction d'importance basée sur la dynamique : SIR
FID2	Fonction d'importance basée sur la dynamique : AUXILIARY
FIM1	Fonction d'importance basée sur la mesure : MSIR
FIM2	Fonction d'importance basée sur la mesure : RBSS
FIDM1	Fonction d'importance basée sur la dynamique et la mesure : MSIR combiné avec SIR
FIDM2	Fonction d'importance basée sur la dynamique et la mesure : RBSS combiné avec SIR

TAB. 4.4 – Tableau récapitulatif des fonctions d'importance.

Modalité n°	Mesures	Algortihmes de filtrage
#1	contours+flot optique ($F1M$)	$FID1$
#2	détections visages+*blobs* « peau » ($FSBD$) contours+distribution de couleur ($F2C1$)	$FIDM2$
#3	détection de mouvement (MD) distributions couleur+mouvement	$FIDM1$ ou $FIDM2$

TAB. 4.5 – Tableau récapitulatif des fonctions visuelles relatives à chaque modalité.

Chapitre 5

Reconnaissance de gestes

Dans le chapitre précédent, nous avons considéré des fonctions visuelles de suivi associées à trois modalités d'interaction entre l'homme et un robot guide de musée. Pour enrichir cette interaction dans notre contexte, nous proposons ici des modalités d'interaction gestuelle par vision. Ce chapitre est un peu particulier, puisqu'il décrit la dernière partie de nos travaux, la moins aboutie. Il se peut qu'il laisse un certain goût d'inachevé, bien qu'il reflète à notre avis des travaux significativement entamés et qui nous semblent prometteurs. Ceux-ci ont contribué à deux de nos publications [Brèthes et al., 2004a, Brèthes et al., 2004b].

5.1 Généralités

Le robot Rackham dispose de diverses interfaces embarquées pour interagir avec les individus (figure 5.1) :

- l'écran tactile permet l'affichage d'informations à destination de l'utilisateur qui peut dialoguer *via* appuie sur l'écran ;
- le haut-parleur et microphone embarqués permettent au robot de communiquer oralement des informations à son interlocuteur ou de dialoguer avec lui par reconnaissance vocale ;
- la caméra EVI-D70, située à l'arrière du robot, lui permet de communiquer avec son interlocuteur par reconnaissance gestuelle.

Tout comme la parole, le geste apparaît pour les humains comme un moyen spontané de communication. Il semble donc légitime d'intégrer des capacités de reconnaissance gestuelle sur les robots actuels.

L'encyclopédie Hachette Multimédia définit le geste comme un « mouvement des mains, des bras ou de la tête, effectué avec ou sans intention de signifier quelque chose ».

Dans [Cadoz, 1994], Cadoz s'intéresse au canal gestuel associé à la main, pour lequel il considère trois fonctions distinctes mais complémentaires intervenant à des degrés différents dans chacune des deux autres : (i) une fonction d'action matérielle, de modification et de transformation de l'environnement, nommée fonction *ergotique*, (ii) une fonction *épistémique* de perception de l'environnement, ainsi que (iii) une fonc-

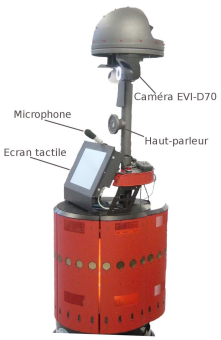


FIG. 5.1 – Interfaces embarquées sur Rackham

tion d'émission d'information à destination de l'environnement dite fonction *sémiotique*. [Quek, 1994] propose une taxonomie des gestes dans laquelle les gestes dit *communicatifs* correspondent à la fonction sémiotique. Cette classification, déclinée sur la figure 5.2, distingue clairement, parmi les gestes communicatifs, les gestes modélisants des gestes référentiels.

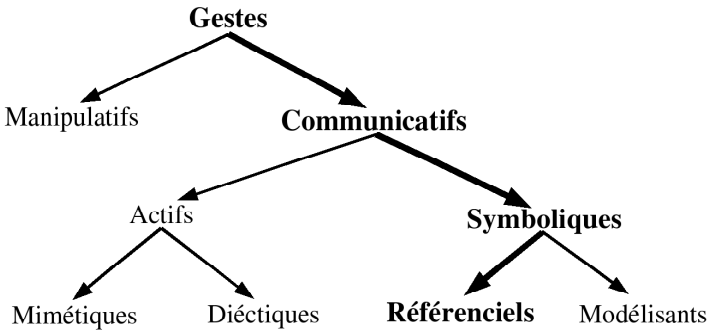


FIG. 5.2 – Taxonomie des gestes proposée par QUEK [Quek, 1994]

Dans notre contexte, nous nous intéressons plus particulièrement aux gestes symboliques de type *référentiels*. Ces gestes correspondent à des messages informationnels et font directement référence à un objet ou un concept, par exemple pour donner un ordre au robot. Considérant uniquement la main, l'information transmise est contenue dans la configuration de la main et/ou le geste effectué. Pavlovic dans [Pavlovic et al., 1997] rappelle que le geste se décompose en trois étapes :

1. la *préparation*, durant laquelle la main est déplacée vers la position de départ du geste ;
2. le *noyau*, qui correspond à la phase de réalisation effective du geste ;
3. la *rétraction*, où la main revient à sa position de départ avant l'exécution éventuelle du geste suivant.

Ces étapes permettent de (i) distinguer les gestes pertinents des mouvements non-intentionnels, (ii) isoler l'information communiquée par les gestes pertinents des deux étapes non-informatives. Les étapes non-informatives ainsi que les mouvements non-intentionnels sont généralement caractérisés par un mouvement rapide de la main alors que le mouvement relatif au noyau est plus lent. Notons également que le noyau peut être plus facilement discerné s'il est délimité par des configurations particulières de la main.

Ayant isolé le noyau du geste, la reconnaissance peut alors s'effectuer. La figure 5.3 montre le synoptique d'un système de reconnaissance classique de gestes.

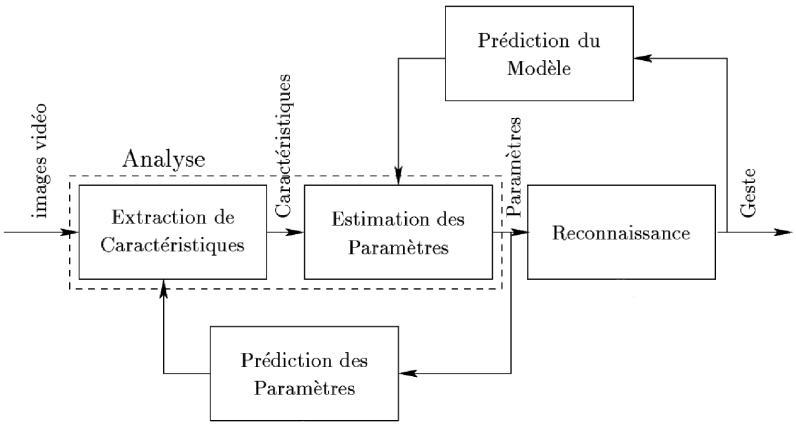


FIG. 5.3 – Synoptique classique d'un système d'analyse/reconnaissance de gestes [Pavlovic et al., 1997]

Le processus est en général séquentiel. Nous distinguons une phase d'analyse, ou « suivi » qui englobe l'extraction des caractéristiques dans l'image courante puis l'estimation des paramètres de la main. La phase de reconnaissance considère alors les paramètres estimés afin d'identifier le geste effectué parmi une base de gestes connus. Un geste « idéal » est donc défini par une trajectoire dans l'espace des paramètres relatifs à son modèle. En pratique, ce geste est classiquement représenté par un nuage de de points autour de cette trajectoire idéale.

Beaucoup de travaux s'inscrivent dans cette démarche. Focalisons nous sur les gestes de la main et dissocions les approches 2D des approches 3D qui sortent du cadre de ces travaux. Les techniques s'appuyant sur un gant numérique [Liang et al., 1995] ou marqueurs [Davis et al., 1994], peu naturelles, sont hors contexte. Concernant la problématique générale, le lecteur trouvera une bibliographie exhaustive dans [Pavlovic et al., 1997] ou plus récente dans [Gavrila, 1999, Wu et al., 1999]. Décrivons plus spécifiquement quelques travaux/techniques de vision monoculaire proches de notre contexte. Nous distinguons deux types de stratégies pour aborder ce problème de reconnaissance :

[A] Une stratégie basée sur une décomposition du geste comme un enchaînement

de positions image. Le geste courant n'est donc pas segmenté au préalable.

[B] Une stratégie basée sur une classification des trajectoires réalisées durant le geste lorsque celui-ci peut être préalablement segmenté. Comparer deux trajectoires revient alors à comparer deux courbes dans l'espace multidimensionnel associé.

Concernant les approches s'inscrivant dans [A], la reconnaissance s'effectue classiquement par réseaux de neurones [Boehm et al., 1994], réseaux dynamiques Bayésiens [Pavlovic et al., 2000] ou plus largement par Modèles de Markov Cachés notés –HMMs : Hidden Markov Models– [Yang et al., 1997, Marcel et al., 2000, Kapuscinski et al., 2001, Yoon et al., 2001]. Ceux-ci sont utilisés par ailleurs pour la reconnaissance de la parole [Rabiner, 1989]. La reconnaissance par HMM repose ici sur la mise à jour, à chaque instant image, de la séquence du vecteur d'état la plus vraisemblable au sens où elle maximise la probabilité d'occurence de la séquence d'observation conjointement à la trajectoire d'état. Mentionnons ici les travaux de Yoon *et al.* dans [Yoon et al., 2001]. Ils concernent les gestes alphabétiques réalisés de façon quasi fronto-parallèle à la caméra. Les attributs visuels position, orientation et vitesse constituent alors les observations pour les HMMs, chaque HMM représentant une lettre de l'alphabet. Ils sont déduits classiquement d'une segmentation de la main à partir d'un histogramme dans l'espace (YIQ) modélisant la couleur peau. Notons que, Kapuscinski *et al.* dans [Kapuscinski et al., 2001] proposent une approche très similaire. Ils se limitent à la reconnaissance de 10 gestes mais incluent la classification (à l'aide d'une transformation morphologique *hit-miss*) de 5 configurations de la main pour dissocier les étapes du geste.

Citons enfin les travaux de Isard et Blake dans [Isard et al., 1998d]. Les auteurs définissent trois modèles de mouvements canoniques apparents caractéristiques d'une main en action de dessiner. L'approche ne dissocie plus aussi distinctement les phases de suivi et de reconnaissance puisque celles-ci s'effectuent simultanément par une variante de la CONDENSATION. Cette variante permet la gestion de variables discrètes indexant les modèles de dynamique et de variables continues relatives à la paramétrisation de l'évolution de la main. Rittscher *et al.* dans [Rittscher et al., 1999] étendent cette approche à des classes de modèles auto-régressifs dont les paramètres sont estimés hors-ligne sur des séquences-test.

Citons quelques travaux s'inscrivant dans la stratégie de type [B]. Ils s'appuient souvent sur les techniques connues de classification et reconnaissance des formes.

Cui *et al.* dans [Cui et al., 1995] segmentent chaque geste sur la base du mouvement inter-images. Le geste est alors représenté par un vecteur concaténant les positions à chaque instant. Les auteurs utilisent alors une Analyse Factorielle Discriminante pour classifier 28 signes de la main extraits du langage proposé par Bornstein *et al.* [Bornstein et al., 1989]. Cette technique est basée sur une partition *a priori* des classes permettant de définir un espace minimisant la distance intra-classe et maximisant la variance interclasse. La règle de décision est caractérisée par une fonction d'interpolation dans cet espace.

Black *et al.* dans [Black et al., 1998] proposent une extension de l'algorithme de CONDENSATION pour la reconnaissance de 6 gestes modélisés *a priori* par leurs trajectoires image dans le flot vidéo. Les composantes du vecteur d'état estimé caractérisent les modèles de trajectoires à reconnaître, *i.e.* un index référençant les modèles des paramètres de déformation des trajectoires associées. Les trajectoires propres aux gestes effectués sont segmentées à partir d'histogrammes couleurs.

Chateau *et al.* dans [Chateau et al., 2004] proposent une méthode de reconnaissance de 18 gestes. Un algorithme de CONDENSATION permet de suivre les deux mains dans l'espace. La reconnaissance des gestes consiste à comparer les trajectoires 2D des poignets dans l'image avec la base des gestes-types. Les auteurs utilisent ici la distance partielle de Hausdorff afin de mesurer la distance entre un sous-ensemble des deux nuages de points associés.

Shan *et al.* dans [Shan et al., 2004] effectuent le suivi de la main par une technique combinant *mean-shift* et filtrage particulaire afin de reconnaître 7 gestes, préalablement représentés par des motifs temporels codant l'historique du mouvement image (*Motion History Image*). Ceux-ci sont caractérisés par des vecteurs englobant les sept moments de Hu. La règle de décision repose alors sur une distance de Mahalanobis entre vecteurs pour chaque modèle et le geste à classifier.

Intéressons-nous à la reconnaissance de configurations associées à la main. Triesch *et al.* dans [Triesch et al., 2002] proposent une technique d'appariement de graphes pour reconnaître 10 configurations de main. La topologie de chaque graphe est représentative d'une configuration donnée : les nœuds du graphe sont labellisés par des descripteurs image locaux basés sur des filtres de Gabor tandis que les arcs représentent les distances image entre nœuds. L'approche proposée est robuste aux arrières-plans encombrés, invariante à l'échelle mais non au point de vue car un seul graphe est défini par posture. Notons enfin que l'approche ne requiert pas de segmentation préalable de l'image.

Thayananthan *et al.* dans [Thayananthan et al., 2003] modélisent un ensemble de configurations par leurs silhouettes. Celles-ci sont déclinées suivant un arbre afin de faciliter la reconnaissance de la configuration courante. Cet arbre est construit hors-ligne par une technique de *K-means*. Pour la reconnaissance, la fonction de vraisemblance repose sur des critères combinant :

- la forme grâce à une distance de Chanfrein sur les contours extraits,
- la couleur à l'intérieur [resp. extérieur] de la silhouette grâce à la distribution de couleur peau [resp. fond].

Bretzner *et al.* dans [Bretzner et al., 2002] définissent 5 configurations de la main à suivre et reconnaître dans le flot vidéo. La main est caractérisée par un ensemble hiérarchique d'ellipses et de cercles pour, d'une part les doigts, d'autre part leurs extrémités et la paume. Les caractéristiques image associées sont détectées dans l'image grâce aux invariants différentiels proposés par Lindeberg [Lindeberg, 1998]. Le suivi et la reconnaissance de la configuration s'effectuent simultanément par une variante de la CONDENSATION avec échantillonnage hiérarchique. La fonction de mesure est définie par une différence quadratique entre mélanges de Gaussiennes relatives aux caractéristiques modèle et image.

Liu *et al.* dans [Liu et al., 2004] proposent une approche très similaire. Les auteurs définissent 4 configurations de la main modélisées par leurs silhouettes. La fonction de mesure repose sur le seul critère de forme, ce qui laisse penser que l'approche est peu robuste aux scènes encombrées.

Comme [Isard et al., 1998d, Triesch et al., 2002, Bretzner et al., 2002, Liu et al., 2004], notre approche pour la reconnaissance de gestes ne dissocie plus aussi distinctement les deux phases d'analyse et de reconnaissance de la figure 5.3. Il s'agit ici d'effectuer simultanément l'estimation et la reconnaissance de configurations et/ou de mouvements image « bas-niveau » dans le processus d'analyse de la main. Certaines stratégies de filtrage particulaire, présentées ci-après, nous semblent ici tout indiquées. Elles permettent d'estimer simultanément :

- les paramétres continus relatifs à la position image courante,
- les paramètres discrets indexant des configurations et/ou de modèles canoniques de dynamique de la main suivie.

Cette stratégie de classification en modèles de dynamiques peut être assimilée à une segmentation de la séquence analysée [Rittscher et al., 1999]. Elle n'exclut pas un processus final de reconnaissance de gestes, en particulier pour des gestes plus complexes, *e.g.* incluant une sémantique (voir § 5.5). Un geste sera alors caractérisable par un séquencement de trajectoires canoniques *segmentées* dans le flot vidéo. Notre démarche, située à mi-chemin entre les stratégies [A] et [B] précédemment décrites, permet une représentation plus compacte des gestes à reconnaître. Nous espérons ainsi gagner en simplicité de mise en œuvre.

Le plan du chapitre est le suivant. Le § 5.2 décrit deux stratégies de filtrage particulaire permettant l'estimation conjointe de paramètres continus et discrets sous l'hypothèse que le problème puisse être modélisé dans le cadre des systèmes à sauts Markoviens. Le § 5.3 reprend le scenario initial de notre robot-guide et spécifie nos modalités d'interaction gestuelle dans ce contexte. Enfin, les sections 5.4 et 5.5 décrivent les fonctions visuelles associées à ces modalités. Leurs implémentations et les résultats associés obtenus ou attendus sont alors présentés et discutés.

5.2 Filtrage particulaire et systèmes dynamiques à sauts Markoviens

5.2.1 Généralités

Comme indiqué en introduction du §2, le filtrage particulaire permet l'estimation du vecteur d'état de tout système dynamique, que celui-ci soit continu, discret, ou hybride. Ainsi, soit $X_{0:k} = (r_{0:k}, x_{0:k})$ le processus aléatoire d'état jusqu'à l'instant k, avec $r_{0:k}$ et $x_{0:k}$ les sous-processus à valeurs discrètes et continues, respectivement. Disposant de la réalisation $z_{1:k}$ du processus de mesure, le but est donc d'estimer la densité *a posteriori*

$p(X_{0:k}|z_{1:k})$ au moyen de l'approximation particulaire[1]

$$p(X_{0:k}|z_{1:k}) = p(r_{0:k}, x_{0:k}|z_{1:k}) \approx \sum_{i=1}^{N} w_{0:k}^{(i)} \delta_{r_{0:k}, r_{0:k}^{(i)}} \delta(x_{0:k} - x_{0:k}^{(i)}), \tag{5.1}$$

ou bien la densité de filtrage, avec $w_k^{(i)} = w_{0:k}^{(i)}$,

$$p(X_k|z_{1:k}) = p(r_k, x_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta_{r_k, r_k^{(i)}} \delta(x_k - x_k^{(i)}). \tag{5.2}$$

Outre les hypothèses (2.2),(2.1) – relatives au caractère Markovien de la dynamique et à l'indépendance, conditionnellement au processus d'état, de la mesure courante avec les mesures passées et avec l'état suivant – qui permettent de décrire entièrement tout système considéré au moyen de sa loi de dynamique $p(X_k|X_{k-1})$ et de son lien état-mesure $p(z_k|X_k)$, on admet également que le processus $\{r_k\}_{k\in\mathbb{N}^*}$ se réalise dans un ensemble discret fini S selon une chaîne de Markov du premier ordre, homogène, stationnaire. En d'autres termes, l'étude est limitée aux systèmes dynamiques à sauts Markoviens – JMS : *Jump-Markov Systems* – dont la loi de dynamique satisfait par hypothèse

$$\begin{aligned} p(X_k|X_{0:k-1}) &= p(X_k|X_{k-1}) \\ &= p(x_k|r_{k-1}, x_{k-1}, r_k) p(r_k|r_{k-1}, x_{k-1}), \\ &= p(x_k|r_{k-1}, x_{k-1}, r_k) p(r_k|r_{k-1}), \end{aligned} \tag{5.3}$$

et qui sont entièrement décrits par la distribution *a priori* $p(x_0)$, les probabilités de transition des états discrets – stationnaires et indépendantes du processus d'état continu –

$$\pi_{ij} \triangleq P(r_k = j|r_{k-1} = i), \ \forall k \geq 1, \ \forall (i,j) \in S \times S, \tag{5.4}$$

les dynamiques continues

$$p_{r_{k-1} r_k}(x_k|x_{k-1}) \triangleq p(x_k|r_{k-1}, x_{k-1}, r_k) \tag{5.5}$$

et les liens état-mesure

$$p_{r_k}(z_k|x_k) \triangleq p(z_k|x_k, r_k). \tag{5.6}$$

De nombreux phénomènes peuvent être modélisés en tant que systèmes à sauts Markoviens, tels les gestes. Supposons en effet qu'un geste consiste en l'enchaînement de gestes élémentaires « canoniques », chacun d'eux étant indicé par l'un des s couples (configuration,dynamique) constituant une « bibliothèque » $S = \{1, \ldots, s\}$ définie *a priori*. À l'instant k, soient $r_k \in S$ l'indice du geste canonique en cours d'exécution et

[1] Dans tout le chapitre, où r [resp. x] désigne une variable aléatoire discrète [resp. continue], on commet l'abus de notation consistant à désigner par $\delta(x - x^{(i)})$ la densité de probabilité définie sur l'espace continu $x \in \mathbb{R}^{n_x}$ et par $\delta_{r,r^{(i)}}$ le nombre $\mathbb{P}(r = r^{(i)})$, égal à 1 [resp. 0] ssi $r = r^{(i)}$ [resp. $r \neq r^{(i)}$].

x_k le vecteur relatif à la paramétrisation du mouvement de la main – position, orientation, vitesses, etc. – à l'intérieur de ce geste élémentaire. Une telle formalisation s'inscrit dans le cadre des systèmes à sauts Markoviens dès lors qu'on admet que les évolutions possibles entre éléments de la bibliothèque S peuvent être définies *a priori*, indépendamment des valeurs prises par x_k.

Notons qu'une grande flexibilité est permise dans la définition du vecteur d'état continu x_k. En effet, le nombre et la sémantique de ses composantes peut varier significativement selon la valeur de r_k. En outre, la cohérence des valeurs admissibles de ce vecteur en deux instants consécutifs $k-1$ et k peut être capturée au moyen de (5.5) y compris si ces deux instants sont séparés par un saut, *i.e.* si $r_{k-1} \neq r_k$.

La hiérarchie (5.3) existant entre les composantes discrètes et continues du vecteur d'état d'un système à sauts Markoviens permet la définition d'algorithmes de filtrage particulaire simplifiés, où les parties discrètes et continues des particules peuvent être échantillonnées successivement, cf. §5.2.2. Le §5.2.3 introduit une extension, développée dans la littérature, mettant à profit cette hiérarchie de façon à améliorer l'efficacité du filtre.

5.2.2 L'algorithme "mixed-state CONDENSATION"

La CONDENSATION a été adaptée aux système à sauts Markoviens dans [Isard et al., 1998c], donnant ainsi lieu à l'algorithme "mixed-state CONDENSATION". Celui-ci possède une structure semblable à celle de la Table 2.5 page 37. Or, la dynamique selon laquelle la particule $X_k^{(i)} = (r_k^{(i)}, x_k^{(i)})$ est échantillonnée s'écrit, d'après (5.3),

$$p(X_k|X_{k-1}^{(i)}) = \pi_{r_{k-1}^{(i)} r_k} p_{r_{k-1}^{(i)} r_k}(x_k|x_{k-1}^{(i)}). \tag{5.7}$$

Cette expression suggère de sélectionner d'abord l'indice discret $r_k^{(i)}$ dans S selon les probabilités de transition $p(r_k|r_{k-1}^{(i)}) = \pi_{r_{k-1}^{(i)} r_k}$, préalablement à l'échantillonnage de $x_k^{(i)}$ selon $p_{r_{k-1}^{(i)} r_k^{(i)}}(x_k|x_{k-1}^{(i)})$. Par ailleurs, l'étape de calcul des poids (item 7 de l'algorithme Table 2.5) repose sur l'évaluation de la vraisemblance $p(z_k|X_k^{(i)}) = p_{r_k^{(i)}}(z_k|x_k^{(i)})$.

L'estimé du maximum *a posteriori* de r_k, *i.e.* $[\hat{r}_k]_{\text{MAP}} = \arg\max_{r_k} p(r_k|z_{1:k})$, peut être approximé par

$$\hat{r}_k = \arg\max_l \sum_{i \in \Upsilon_l} w_k^{(i)}, \text{ avec } \Upsilon_l = \{i \,:\, X_k^{(i)} = (l, x_k^{(i)})\}, \tag{5.8}$$

et il vient ensuite

$$\hat{x}_k = \frac{\sum_{i \in \Upsilon_{\hat{r}_k}} w_k^{(i)} x_k^{(i)}}{\sum_{i \in \Upsilon_{\hat{r}_k}} w_k^{(i)}}, \text{ où } \Upsilon_{\hat{r}_k} = \{i \,:\, X_k^{(i)} = (\hat{r}_k, x_k^{(i)})\}. \tag{5.9}$$

5.2.3 Une stratégie AUXILIARY_UNSCENTED pour l'estimation du vecteur d'état des systèmes à sauts Markoviens

Bien que l'algorithme présenté dans la section précédente fournisse une solution au problème du filtrage pour des systèmes à sauts Markoviens, il souffre des mêmes problèmes d'efficacité que la CONDENSATION dans le cas du filtrage mono-modèle. Les raisons sont identiques, au sens où les particules hybrides $X_k^{(i)}$ sont échantillonnées selon la dynamique, sans garantie d'exploration de zones de l'espace d'état vraisemblables vis à vis de l'observation. Un schéma se rapprochant de la stratégie récursive optimale, proposé dans [Andrieu et al., 2003], permet de contourner ce problème. Cet algorithme que nous décrivons ci-dessous est en cours d'évaluation. L'idée consiste à appliquer la stratégie mixte AUXILIARY_UNSCENTED présentée au §2.3.3, où $x_k^{(i)}$ est remplacé par $X_k^{(i)} = (r_k^{(i)}, x_k^{(i)})$. Du fait que

$$\begin{aligned} p(z_k|X_{k-1}) &= \sum_r p(r|r_{k-1}, x_{k-1}) p(z_k|r, r_{k-1}, x_{k-1}) \\ &= \sum_r \pi_{r_{k-1}r} p(z_k|r, r_{k-1}, x_{k-1}), \end{aligned} \tag{5.10}$$

l'approximation $\hat{p}(z_k|X_{k-1})$ de $p(z_k|X_{k-1})$, qui intervient dans la définition des pondérations auxiliaires sur lesquelles est basé le rééchantillonnage intermédiaire, repose sur les approximations $\hat{p}(z_k|r, r_{k-1}, x_{k-1})$ de $p(z_k|r, r_{k-1}, x_{k-1})$, $r \in S$, au moyen de la transformée unscented. À cette fin, pour chaque particule $X_{k-1}^{(i)}$, un ensemble $\Xi_{k-1}^{(i)}$ de $n_{k-1}^{(i)}$ σ-points est obtenu par échantillonnage déterministe selon la loi Gaussienne $\mathcal{N}(x_{k-1}; m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)})$ qui lui est associée. Les auteurs proposent alors de définir

$$\hat{p}(z_k|r, r_{k-1}^{(i)}, x_{k-1}^{(i)}) = \frac{1}{n_{k-1}^{(i)}} \sum_{\xi_{k-1}^{(i)} \in \Xi_{k-1}^{(i)}} p(z_k|r, \underline{r_{k-1}^{(i)} r}(\xi_{k-1}^{(i)})), \tag{5.11}$$

où $\underline{r_{k-1}^{(i)} r}(\xi)$ désigne l'image de chaque σ-point ξ par la dynamique continue relative à la transition $r_{k-1}^{(i)} \to r$, dans laquelle les termes de bruits sont préalablement fixés à 0. Ensuite, le rééchantillonnage auxiliaire correspondant à l'item 11 de l'algorithme AUXILIARY_UNSCENTED Table 2.11 page 56 convertit l'ensemble des particules pondérées $\{X_{k-1}^{(i)}, \lambda_k^{(i)} \propto w_{k-1}^{(i)} \sum_r \pi_{r_{k-1}^{(i)} r} \hat{p}(z_k|r, r_{k-1}^{(i)}, x_{k-1}^{(i)})\}$ – qui représente le lisseur $p(X_{k-1}|z_{1:k})$ du fait que $\sum_r \pi_{r_{k-1}^{(i)} r} \hat{p}(z_k|r, r_{k-1}^{(i)}, x_{k-1}^{(i)})$ approxime $p(z_k|X_{k-1}^{(i)})$ – et ses statistiques associées $\{m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}\}$, en l'ensemble équipondéré équivalent $\{\tilde{X}_{k-1}^{(i)}, \frac{1}{N}\}$ et les statistiques $\{\tilde{m}_{k-1|k-1}^{(i)}, \tilde{P}_{k-1|k-1}^{(i)}\}$.

Cette méthode de définition des pondérations auxiliaires possède en outre l'avantage de permettre une hiérarchisation de l'échantillonnage de $X_k^{(i)} = (r_k^{(i)}, x_k^{(i)})$. Dans un premier temps, les indices $r_k^{(i)}$ sont sélectionnés dans S selon la distribution

$q(r_k|\tilde{X}_{k-1}^{(i)}, z_k) \propto \pi_{\tilde{r}_{k-1}^{(i)} r_k} \hat{p}(z_k|r_k, \tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)})$ qui mime la fonction d'importance optimale du fait que $\pi_{\tilde{r}_{k-1}^{(i)} r_k} \hat{p}(z_k|r_k, \tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}) = \hat{p}(z_k, r_k|\tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}) \propto \hat{p}(r_k|\tilde{X}_{k-1}^{(i)}, z_k)$.

Ensuite, les moments $m_{k|k}^{(i)}$ et $P_{k|k}^{(i)}$ de la Gaussienne $\mathcal{N}(x_k; m_{k|k}^{(i)}, P_{k|k}^{(i)})$ associée à $x_k^{(i)}$ sont calculés par un pas de l'UKF, pour la dynamique continue relative à la transition $\tilde{r}_{k-1}^{(i)} r_k^{(i)}$[2]. Après que $x_k^{(i)}$ soit échantillonné selon $\mathcal{N}(x_k; m_{k|k}^{(i)}, P_{k|k}^{(i)})$, les poids sont mis à jour de façon à obtenir une approximation particulaire cohérente de la densité de filtrage telle qu'en (5.2).

L'algorithme est résumé Figure 5.1.

5.3 Scénario et modalités d'interaction gestuelle associées

L'interaction gestuelle s'intègre dans les modalités d'interaction associées au scenario décrit au § 4 et plus globalement dans le cadre de l'interaction entre un robot guide et des visiteurs. Lors de la phase #2 du scenario, l'interlocuteur du robot sélectionne sur l'écran tactile un lieu de l'exposition vers lequel il souhaite être guidé par le robot. La phase #3 consiste alors pour le robot à planifier puis exécuter une trajectoire afin de guider le visiteur vers le lieu de son choix. Le robot reste en contact avec son interlocuteur *via* sa caméra « arrière » durant l'exécution de cette mission. L'individu guidé doit, à tout moment, pouvoir interrompre ou modifier le but de la mission. Une interaction à distance, donc par reconnaissance visuelle de gestes, nous semble ici tout indiquée. L'interlocuteur du robot pourra, par exemple, notifier un changement de but par une configuration de la main indexant un nouveau but ou stand (supposé référencé dans l'exposition) sur lesquels le robot doit commuter durant l'exécution de cette mission.

Rappelons que ces travaux s'inscrivent dans une collaboration avec la Cité de l'Espace dont l'objectif premier est de sensibiliser le grand public par un démonstrateur ludique. Dans ce cadre, un jeu est proposé aux interlocuteurs du robot. Nous aimerions en effet voir ceux-ci s'identifier par leurs prénoms en début ou fin de mission. Le robot pourrait alors les interpeller ultérieurement par leurs prénoms grâce au haut-parleur lorsqu'il seraient identifiés dans le flot vidéo par le classifieur de visages (voir § 4.3.2). L'objectif final est de caractériser l'alphabet par un ensemble de configurations de la main, de modèles de dynamique et d'enchaînements entre ceux-ci.

Nos deux modalités d'interaction gestuelle se résument donc comme suit :

1. **Changement de but :** Durant l'exécution de la mission, le visiteur guidé repère un stand qui lui semble tout autant intéressant. Il en informe alors à distance le robot par une configuration de la main référençant ce nouveau but.

[2]Notons que si le bruit de dynamique est additif, cette étape réutilise l'ensemble $\tilde{\Xi}_{k-1}^{(i)}$ des σ-points relatifs à $\mathcal{N}(x_{k-1}; \tilde{m}_{k-1|k-1}^{(i)}, \tilde{P}_{k-1|k-1}^{(i)})$, de même que son image $\underline{\tilde{r}_{k-1}^{(i)} r_k^{(i)}}(\tilde{\Xi}_{k-1}^{(i)})$ *via* la dynamique non bruitée associée à transition $\tilde{r}_{k-1}^{(i)} r_k^{(i)}$ sélectionnée lors de l'échantillonnage – quasi-optimal– de r_k, ce qui limite significativement la complexité calculatoire de l'algorithme.

$[\{r_k^{(i)}, x_k^{(i)}, w_k^{(i)}, m_{k|k}^{(i)}, P_{k|k}^{(i)}\}]_{i=1}^N = \text{JMSPF}([\{r_{k-1}^{(i)}, x_{k-1}^{(i)}, w_{k-1}^{(i)}, m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}\}]_{i=1}^N, z_k)$

SI $k = 0$ **(INITIALISATION) ALORS**
  Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$
  Initialiser les moyennes $m_{0|0}^{(i)}$ et covariances $P_{0|0}^{(i)}$ associées à $x_0^{(i)}$,$i = 1, \ldots, N$
FIN SI
SI $k \geq 1$ **ALORS**
  POUR $i = 1, \ldots, N$, **FAIRE**
    Par utilisation de la transformée unscented, calculer l'approximation $\hat{p}(z_k|r_{k-1}^{(i)}, x_{k-1}^{(i)})$ à partir de $\{m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}\}$ comme suit :
      échantillonner de manière déterministe un ensemble $\Xi_{k-1}^{(i)}$ de $n_{k-1}^{(i)}$ σ-points répartis selon $\mathcal{N}(x_{k-1}; m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)})$
      POUR chaque indice $r \in S$, **FAIRE**
        calculer l'image $\underline{r_{k-1}^{(i)} r}(\Xi_{k-1}^{(i)})$ de l'ensemble $\Xi_{k-1}^{(i)}$ par la dynamique continue associée à la transition $r_{k-1}^{(i)} r$, les termes de bruits étant fixés à 0
        calculer l'approximation $\hat{p}(z_k|r, r_{k-1}^{(i)}, x_{k-1}^{(i)})$ *via* (5.11)
        calculer $\hat{p}(z_k, r|r_{k-1}^{(i)}, x_{k-1}^{(i)}) = \pi_{r_{k-1}^{(i)} r} \hat{p}(z_k|r, r_{k-1}^{(i)}, x_{k-1}^{(i)})$
        calculer $\hat{p}(z_k|r_{k-1}^{(i)}, x_{k-1}^{(i)}) = \sum_r \hat{p}(z_k, r|r_{k-1}^{(i)}, x_{k-1}^{(i)})$
      FIN POUR
    En déduire le poids auxiliaire $\lambda_k^{(i)} \propto w_{k-1}^{(i)} \hat{p}(z_k|r_{k-1}^{(i)}, x_{k-1}^{(i)})$
  FIN POUR
  Rééchantillonner $\{(r_{k-1}^{(i)}, x_{k-1}^{(i)}), \lambda_k^{(i)}\}$ ainsi que les moments et ensembles de σ-points associés $\{m_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)}, \Xi_{k-1}^{(i)}\}$ de façon à obtenir l'ensemble équivalent de particules équipondérées $\{(\tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}), \frac{1}{N}\}$ et les $\{\tilde{m}_{k-1|k-1}^{(i)}, \tilde{P}_{k-1|k-1}^{(i)}, \tilde{\Xi}_{k-1}^{(i)}\}$ associés ; $\sum_{i=1}^N \lambda_k^{(i)} \delta_{r_{k-1}, r_{k-1}^{(i)}} \delta(x_{k-1} - x_{k-1}^{(i)})$ et $\sum_{i=1}^N \frac{1}{N} \delta_{\tilde{r}_{k-1}, \tilde{r}_{k-1}^{(i)}} \delta(\tilde{x}_{k-1} - \tilde{x}_{k-1}^{(i)})$ représentent le lisseur $p(r_{k-1}, x_{k-1}|z_{1:k})$
  POUR $i = 1, \ldots, N$, **FAIRE**
    Échantillonner successivement $r_k^{(i)}$ et $x_k^{(i)}$ comme suit :
      sélectionner $r_k^{(i)} \sim q(r_k|\tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}, z_k) = \frac{\pi_{\tilde{r}_{k-1}^{(i)} r_k} \hat{p}(z_k|r_k, \tilde{r}_{k-1}^{(i)} \tilde{x}_{k-1}^{(i)})}{\sum_r \pi_{\tilde{r}_{k-1}^{(i)} r} \hat{p}(z_k|r, \tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)})}$
      calculer $\{m_{k|k}^{(i)}, P_{k|k}^{(i)}\}$ à partir de $\underline{\tilde{r}_{k-1}^{(i)} r_k^{(i)}}(\tilde{\Xi}_{k-1}^{(i)})$ – l'ensemble $\tilde{\Xi}_{k-1}^{(i)}$ associé à $\{\tilde{m}_{k-1|k-1}^{(i)}, \tilde{P}_{k-1|k-1}^{(i)}\}$ ainsi que son image $\underline{\tilde{r}_{k-1}^{(i)} r_k^{(i)}}(\tilde{\Xi}_{k-1}^{(i)})$ ayant été déja calculés aux items 8,10 – au moyen d'un pas d'UKF, d'abord en tenant compte de la dynamique relative à la transition $\tilde{r}_{k-1}^{(i)} r_k^{(i)}$, puis en incorporant la mesure z_k
      échantillonner $x_k^{(i)} \sim \mathcal{N}(x_k; m_{k|k}^{(i)}, P_{k|k}^{(i)})$
    Mettre à jour les poids, préalablement à leur normalisation, en posant $w_k^{(i)} \propto \frac{p(z_k|r_k^{(i)}, x_k^{(i)}) p(r_k^{(i)}, x_k^{(i)}|\tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)})}{\hat{p}(z_k|\tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}) q(r_k^{(i)}, x_k^{(i)}|\tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}, z_k)}$, ou, de manière équivalente, $w_k^{(i)} \propto \frac{p_{r_k^{(i)}}(z_k|x_k^{(i)}) p_{\tilde{r}_{k-1}^{(i)} r_k^{(i)}}(x_k^{(i)}|\tilde{x}_{k-1}^{(i)})}{\hat{p}(z_k|\tilde{r}_k^{(i)}, \tilde{r}_{k-1}^{(i)}, \tilde{x}_{k-1}^{(i)}) \mathcal{N}(x_k^{(i)}; m_{k|k}^{(i)}, P_{k|k}^{(i)})}$ de sorte que $p(r_k, x_k|z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta_{r_k, r_k^{(i)}} \delta(x_k - x_k^{(i)})$
  FIN POUR
FIN SI

TAB. 5.1 – Stratégie AUXILIARY_UNSCENTED pour les systèmes à sauts Markoviens (JMSPF : *Jump-Markov Systems Particle Filter*)

2. **L'apprentissage de prénoms :** Le robot à l'arrêt souhaite enregistrer le prénom de son interlocuteur. Le protocole d'interaction est spécifié *via* l'écran tactile durant l'interaction proximale (modalité #1). Comme dans [Liang et al., 1995] avec l'alphabet américain, la synthèse vocale reproduit une à une les lettres apprises pour vérification.

5.4 Interaction gestuelle pour le changement de but

5.4.1 Considérations générales

Le but est de modifier la mission en cours en spécifiant un nouveau but (lieu) par le numéro identifiant ce dernier. L'interaction gestuelle repose donc ici sur la reconnaissance de configurations de la main. Celles-ci sont modélisées par leurs silhouettes 2D rigides, que nous appellerons *templates*, déclinées figure 5.4.

FIG. 5.4 – Liste des configurations de la main pour le changement de but (a)-(f), configuration de contrôle (g).

Chaque configuration de la main (figure 5.4-(a - f)) indexe un chiffre donné tandis que l'enchaînement de plusieurs configurations permet la composition de nombres à deux chiffres. La phase de préparation du geste est notifiée par l'enchaînement des configurations « main ouverte » (figure 5.4-(f)) et de contrôle (figure 5.4-(g)).

La première permet d'initialiser le suivi par la configuration la plus discriminante et naturelle ici alors que la deuxième signale le début d'une séquence de configurations. La phase de rétraction est indiquée par la configuration de contrôle, suivie de la configuration « main ouverte » (figure 5.4-(f)) qui confirme la fin de la séquence. La figure 5.5 montre l'enchaînement des configurations permettant d'indiquer au robot de se diriger vers le but 12 de l'exposition.

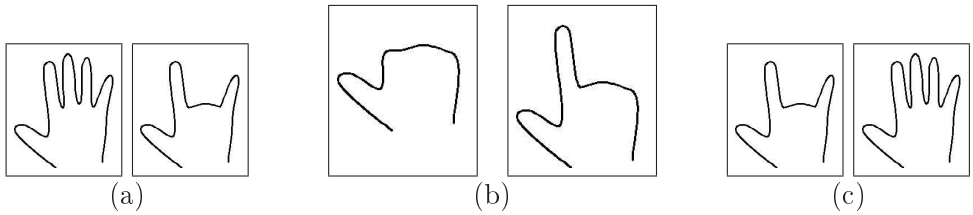


FIG. 5.5 – Enchaînnement indiquant au robot de se diriger vers le but 12 de l'exposition. (a) préparation, (b) noyau et (c) rétraction

Pour cette modalité, le vecteur d'état x_k du filtre doit contenir quatre composantes continues liées respectivement à la position (u_k, v_k), l'orientation θ_k et l'échelle s_k du *template* ainsi qu'un paramètre discret c_k indexant la configuration de la main. Concernant la dynamique du *template*, nous optons pour une marche aléatoire (§ 4.2) qui nous semble caractériser le mieux les mouvements apparents de la cible du fait que la dynamique du geste ne peut donner lieu à interprétation dans cette modalité de guidage. Le vecteur d'état du filtre est alors défini par :

$$X_k = (x_k, c_k)' \text{ avec } x_k = [u_k, v_k, \theta_k, s_k]' \text{ et } c_k \in \{a, b, \ldots, g\}.$$

Les paramètres continus de x_k évoluent ici suivant des marches aléatoires indépendantes. Le paramètre discret noté c_k, indexant les configurations, évolue selon une matrice de transition décrivant les probabilités de passage d'un état discret vers un autre, *i.e.* les commutations entre configurations. Typiquement on prend une matrice de transition dans laquelle le paramètre discrêt à une forte probabilité de rester dans le même état par rapport aux probabilités de changer pour un autre état. Lorsqu'un language est défini, ces probabilités de transitions peuvent alors être afinées par apprentissage.

5.4.2 Fonction de mesure envisagée

L'objectif est ici de suivre la main dans le flot vidéo et de reconnaître sa configuration. Notre stratégie multi-attributs nous amène à fusionner la forme et la couleur dont la distribution est calculée dans des sous-régions de la main. Celle-ci est alors découpée en six sous-régions $\{R_i\}_{i=0,\ldots,5}$ pour lesquelles une distribution de couleur peau est attendue ou pas selon la configuration. Enfin, une septième distribution de couleur est relative à la sous-région R_6 (supposée non peau) définie par les pixels inclus dans le rectangle englobant le *template* mais à l'extérieur de celui-ci. La figure 5.6 montre le découpage du *template* en sous-régions.

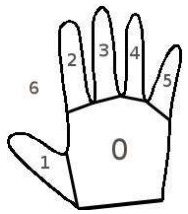


FIG. 5.6 – Découpage de la main en sous-régions

La vraisemblance globale $p(z^{C_{rgb}}|X_k)$ d'une configuration de la main est donnée par le produit des vraisemblances de chaque sous-région. Celles-ci sont calculées relativement à la distribution de couleur peau ou non peau selon la configuration envisagée. Ainsi, pour la configuration illustrée par la figure 5.4-(c) nous avons :

$$p(z^{C_{rgb}}|X_k) = p_0^P . p_1^P . p_2^P . p_3^F . p_4^F . p_5^F . p_6^F$$

où p_i^P et p_i^F sont les vraisemblances relatives aux modèles de couleur peau et fond pour la sous-région R_i considérée.

5.4.3 Évaluation

Le suivi de la main et la reconnaissance de sa configuration parmi les modèles de la figure 5.4 sont évalués sur 20 séquences d'environ 400 images représentatives de l'environnement. La figure 5.7 montre quelques exemples d'images d'interaction pour le changement de but.

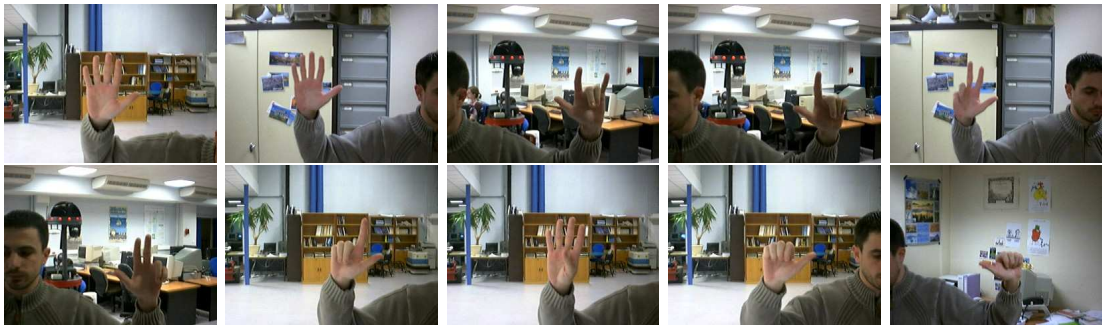


FIG. 5.7 – Exemples d'images acquises au cours de l'interaction

Pour chaque séquence traitée, des taux de reconnaissance moyens et pour chaque configuration sont calculés à partir de plusieurs réalisations de filtrage pour l'algorithme de *mixed-state CONDENSATION*. Le nombre de particules N considéré varie de 50 à 400. Les tests considèrent une fonction de mesure basée sur le seul attribut forme ou la fusion forme et couleur. La figure 5.8 illustre une réalisation de suivi/reconnaissance sur une séquence pour une scène peu encombrée.

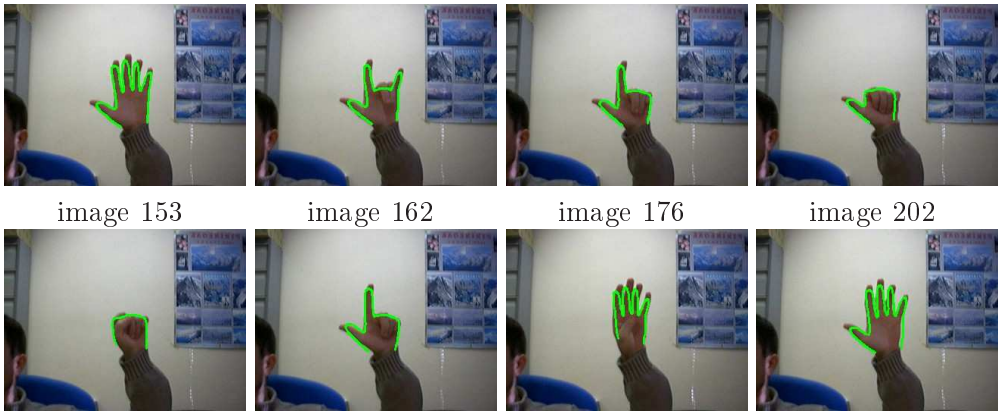


FIG. 5.8 – Réalisation d'un suivi pour un fond peu encombré

Le Tableau 5.4.3 liste les taux de reconnaissance obtenus.

La fonction de mesure fusionnant forme et distribution de couleur permet d'atteindre des taux de reconnaissance moyens élevés (de l'ordre de 97%). Nous constatons que l'utilisation du seul attribut forme dans la fonction de mesure ne permet pas de différencier correctement des configurations qui ne diffèrent que d'un doigt *e.g.* la main ouverte est confondue avec les quatre doigts ouverts, le pouce levé est assimilé à la main fermée,... L'utilisation de la distribution de couleur permet logiquement de mieux différencier ces configurations.

Pour une scène encombrée (figure 5.9), la fusion de plusieurs attributs prend tout son sens. La figure 5.9 montre une réalisation de suivi/reconnaissance dans ce contexte

	Mesure Forme seule					Mesure Forme et Couleur				
	N=50	N=100	N=150	N=200	N=400	N=50	N=100	N=150	N=200	N=400
	97.14%	100%	100%	100%	100%	88.57%	91.43%	94.29%	88.57%	97.14%
	15%	20%	5%	15%	37.50%	100%	100%	100%	100%	100%
	62.39%	65.49%	62.83%	76.99%	79.20%	94.25%	99.56%	98.67%	99.56%	99.56%
	85.59%	96.85%	97.75%	94.14%	95.50%	87.39%	93.24%	97.75%	95.95%	99.56%
	100%	100%	100%	100%	100%	96.36%	100%	100%	100%	100%
	5.91%	0.38%	0.79%	0%	0.79%	99.60%	99.23%	96.20%	99.62%	99.60%
	51.14%	60.23%	67.61%	69.32%	59.66%	97.73%	82.39%	97.16%	96.02%	98.30%
Total	**55.13%**	**58.01%**	**59.04%**	**61.02%**	**61.96%**	**94.91%**	**95.13%**	**97.67%**	**97.95%**	**98.68%**

TAB. 5.2 – Taux de reconnaissance moyens et par configuration *vs* nombre de particules pour nos deux fonctions de mesures considérées (scènes peu encombrées).

difficile. Nous observons que la stratégie de fusion permet ici encore de mieux discerner les configurations.

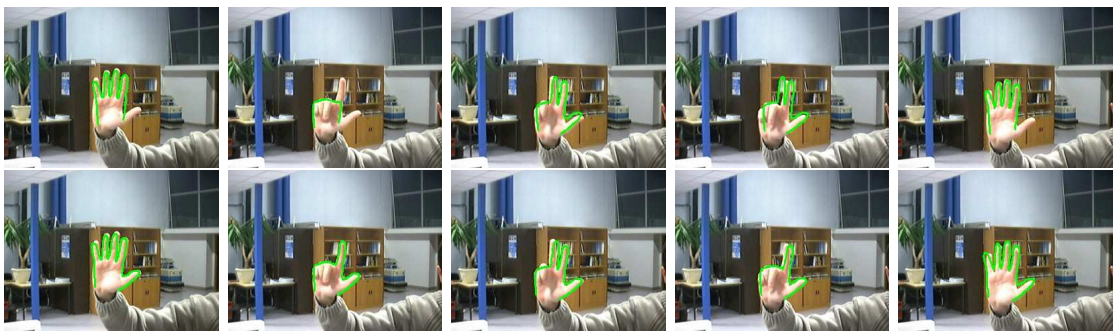


FIG. 5.9 – Exemple de réalisation de suivi de la main pour une scène encombrée avec une mesure basée sur : forme (en haut), forme et distribution de couleur (en bas)

Les taux de reconnaissance obtenus sur les séquences de ce type sont listés dans le tableau 5.3. Les taux de reconnaissance sont logiquement dégradés pour une mesure considérant seulement la forme. La stratégie multi-attributs reste moins perturbée par la complexité de la scène et les taux obtenus restent satisfaisants.

	Mesure Forme seule					Mesure Forme et Couleur				
	N=50	N=100	N=150	N=200	N=400	N=50	N=100	N=150	N=200	N=400
	61.11%	61.11%	66.67%	83.33%	83.33%	100%	94.44%	100%	94.44%	94.44%
	0%	0%	0%	0%	0%	100%	100%	100%	100%	100%
	21.67%	8.33%	13.33%	30%	16.67%	76.67%	75%	73.33%	80%	83.33%
	0%	41.30%	43.30%	43.48%	43.48%	50%	69.57%	89.13%	95.65%	95.65%
	100%	100%	100%	100%	100%	94.44%	100%	94.44%	100%	94.44%
	1.39%	0.92%	4.19%	0%	7.43%	84.79%	95.35%	92.56%	94.91%	95.59%
	11.17%	0%	0%	0%	0%	58.82%	85.29%	94.12%	97.06%	97.06%
Total	**11.17%**	**13.58%**	**17.41%**	**18.30%**	**19.30%**	**78.96%**	**88.81%**	**90.05%**	**93.30%**	**93.86%**

TAB. 5.3 – Taux de reconnaissance moyens et par configuration *vs* nombre de particules pour nos deux fonctions de mesures considérées (scènes encombrées).

5.5 Interaction gestuelle pour l'apprentissage du prénom

5.5.1 Considérations générales

La démarche est ici d'enregistrer le prénom de l'interlocuteur du robot dans la perspective de l'interpeller ultérieurement. Dans ce contexte d'interaction proximale, le robot est arrêté tandis que l'on s'intéresse aux caractéristiques spatio-temporelles du geste afin de reconnaître les lettres qui composent le prénom. Elles sont décrites par des trajectoires inspirées de l'écriture Graffiti développée par Palm Pilot. Ces trajectoires représentées dans la figure 5.10 sont proposées à l'utilisateur qui désire enregistrer son prénom. Le point matérialise pour chaque lettre le départ du geste à réaliser.

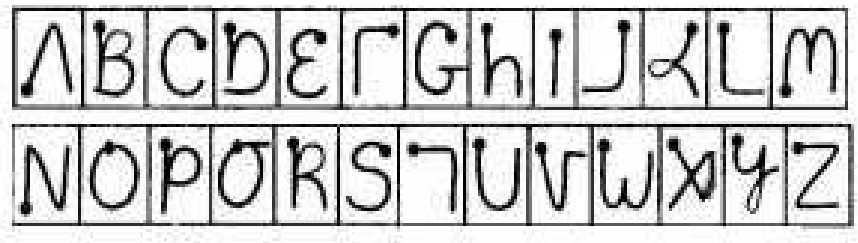


FIG. 5.10 – Écriture Graffiti (Palm) des lettres de l'alphabet

Dans notre approche, les lettres caractérisées par la trajectoire du geste sont représentées par un enchaînement de modèles de dynamiques canoniques estimés dans le

filtrage. Ainsi, la lettre « L », par exemple, est représentée par une dynamique verticale vers le bas suivie d'une dynamique horizontale vers la droite. Il est important de noter que cette représentation nécéssite de pouvoir différencier la direction du mouvement afin par exemple de ne pas confondre « J » avec « L » qui ne diffèrent que par la direction de la dynamique horizontale. La reconnaissance de l'ensemble des lettres de l'alphabet (figure 5.10) implique de définir 10 modèles de dynamique :

- D_E : mouvement horizontal vers la droite,
- D_O : mouvement horizontal vers la gauche,
- D_N : mouvement vertical vers le haut,
- D_S : mouvement vertical vers le bas,
- D_{NE} : mouvement diagonal en haut à droite,
- D_{SE} : mouvement diagonal en bas à droite,
- D_{NO} : mouvement diagonal en haut à gauche,
- D_{SO} : mouvement diagonal en bas à gauche,
- D_C : mouvement circulaire,
- D_I : position immobile.

Le modèle statique (D_I) *a priori* non indispensable est ajouté afin de capturer les pauses pouvant être observées lors de la commutation entre modèles.

Nous utilisons donc une représentation etat/vitesse des paramètres de position de la main pour définir les différents modèles de dynamique. Dans ce contexte, l'orientation et l'échelle de la main ne suivent pas de dynamique particulière et évoluent peu, nous utilisons alors une dynamique de type marche aléatoire pour ces paramètres.

Deux configurations permettent de séparer les phases de préparation et de rétractation du geste. Ainsi, la main ouverte signale l'initialisation et la fin du geste alors que la configuration pouce et index levés (figure 5.4-(c)) marque le noyau du geste effectué.

Le vecteur d'état X_k pour cette modalité inclut donc les mêmes paramètres continus x_k que précédemment ainsi que les dérivées de la position et deux paramètres discrets c_k et d_k indexant respectivement la configuration de la main et son modèle de dynamique. Le vecteur d'état du filtre est alors défini par :

$$X_k = (x_k, y_k) \text{ avec } x_k = [u_k, v_k, \theta_k, s_k, \dot{u_k}, \dot{v_k}]' \text{ et } y_k = [c_k, d_k]$$

$$\text{où } c_k \in \{c, f\}, d_k \in \{D_E, D_O, \ldots, D_I\}$$

et k indexe le temps image dans le flot vidéo.

Comme précédement, les paramètres discrets du vecteur d'état évoluent selon des matrices de transition caractérisant les probabilités de commutation entre les différents modèles. Dans notre cas, ces transitions sont considérées équiprobables. Cependant, une étude détaillée de l'évolution des modèles de dynamique pour les différentes lettres devrait permettre d'adapter les probabilités de transitions, en remarquant par exemple qu'une dynamique horizontale est majoritairement suivie d'une dynamique verticale.

5.5.2 Fonction de mesure envisagée

La fonction de mesure envisagée dans le contexte d'apprentissage de prénom reprend celle utilisée précédemment § 5.4 que nous fusionnons avec une mesure de la distribution

de mouvement. En effet, la mesure vue dans § 5.4 permet de mieux différencier les configurations de la main ainsi que d'éviter un décrochage de la cible tandis que la distribution du mouvement assure un meilleur suivi des trajectoires de la main.

5.5.3 Évaluation

Nous présentons ici une évaluation préliminaire basée sur le calcul des taux de reconnaissance des modèles de dynamiques D_I, D_E, D_O, D_N et D_S. Des travaux sont en cours pour intégrer les autres modèles et les évaluer. Ces cinq premiers modèles permettent néanmoins de reconnaître 5 lettres de l'alphabet. La figure 5.11 montre le résultat d'une réalisation de suivi avec reconnaissance de la configuration de la main et des modèles de dynamiques.

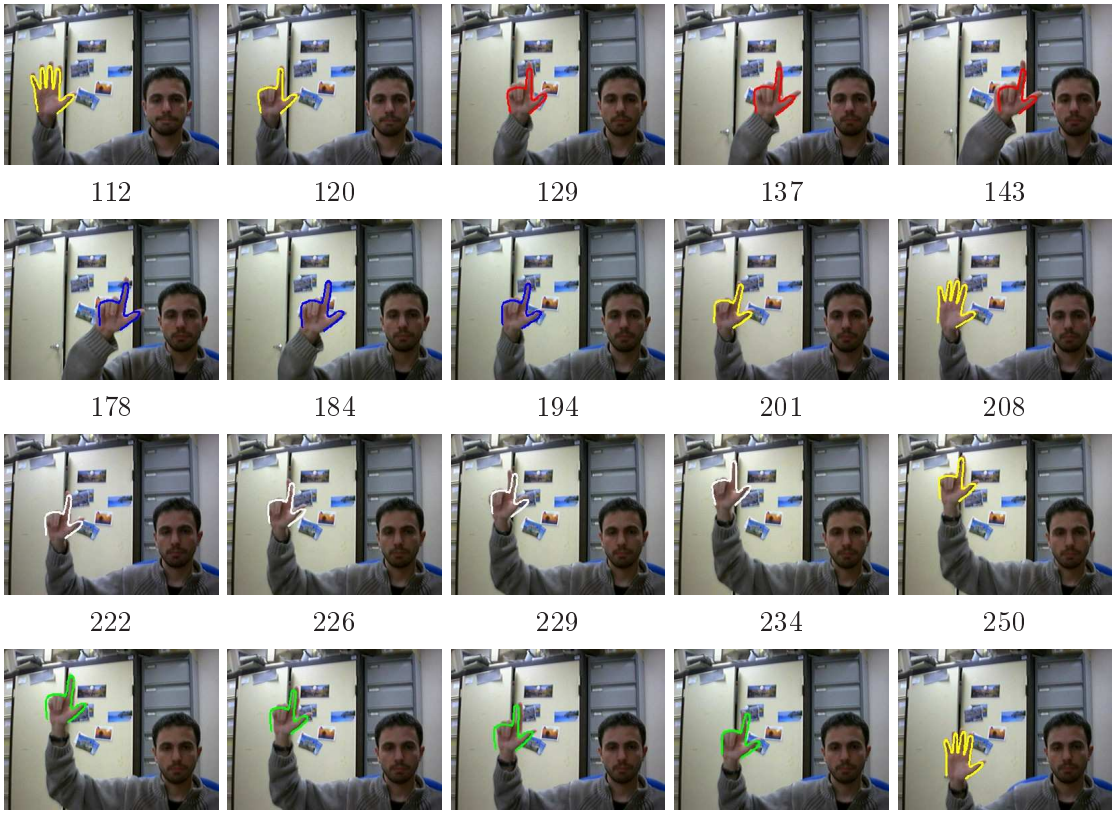


FIG. 5.11 – Exemple de suivi avec reconnaissance des configurations de la main et des modèles de dynamiques D_I (en jaune), D_O (en bleu), D_E (en rouge), D_N (en blanc) et D_S (en vert)

Le tableau 5.4 résume les taux de reconnaissance obtenus à partir 20 réalisations de filtrage appliquées sur des séquences avec fond variable où les 5 lettres ont été réalisées 4 fois chacunes.

	N=50	N=100	N=150	N=200	N=400
D_I	79.37%	87.72%	88.79%	88.70%	90.09%
D_O	87.10%	83.87%	89.25%	86.02%	93.55%
D_E	85.14%	89.19%	82.43%	89.19%	93.24%
D_N	84.31%	86.27%	84.31%	90.20%	90.20%
D_S	81.94%	87.50%	90.28%	93.06%	93.06%
Total	**82.62**%	**86.99**%	**87.43**%	**89.14**%	**91.45**%

TAB. 5.4 – Taux de reconnaissance moyens et par modèle de dynamique *vs* nombre de particules

Les taux de reconnaissance obtenus sont relativements élevés et similaires quelque soit le modèle de dynamique. En pratique, on constate que les erreurs de reconnaissance apparaissent lors de la commutation entre deux modèles et notament lorsque le mouvement entre deux images est peu significatif.

5.6 Conclusion

Dans ce chapitre, nous avons défini deux modalités d'interaction gestuelle entre l'homme et notre « robot-guide ».

La première modalité repose sur la reconnaissance de gestes statiques *i.e.* des configurations de la main pour donner des ordres au robot lors de la mission de guidage. La reconnaissance de la configuration courante et la commutation éventuelle entre configurations sont réalisées dans la boucle de suivi mettant en œuvre un algorithme de *Mixed-state CONDENSATION*. En présence de scènes encombrées, notre fonction de mesure fusionnant les attributs forme et couleur reste discriminante et les taux de reconnaissance sont faiblement dégradés. Une extension envisagée est de prendre en compte des mouvements non fronto-parallèles à la caméra grâce à l'aide d'une transformation affine. Sous hypothèse d'un modèle perspectif faible, nous pourrons alors inférer la position et orientation de la main dans l'espace [Blake et al., 1998b].

La seconde modalité consiste à apprendre et reconnaître des gestes dynamiques représentant l'alphabet afin d'interpeller l'interlocuteur du robot par son prénom. Ces gestes alphabetiques sont modélisés par un séquencement de trajectoires canoniques segmentées dans le flot vidéo. Nous pensons que cette représentation compacte peut simplifier la phase d'apprentissage qui requiert souvent un nombre conséquent de bases de gestes.

Les premières évaluations aboutissent à des taux encourageants de modèles dynamiques. La reconnaissance du modèle de dynamique et la commutation éventuelle entre modèles sont réalisées, ici encore, dans la boucle de suivi à l'aide d'un algorithme de *Mixed-state CONDENSATION*. Nous avons présenté et implémenté un nouvel algorithme de filtrage particulaire (§ 5.2.3) afin de gérer plus finement les différents modèles de dynamique. Cet algorithme relativement original est en cours d'évaluation.

Enfin, concernant la reconnaissance finale de gestes alphabetiques, nous collaborons actuellement avec M.Fox et G.Infantes. Leurs travaux, menés au sein du groupe RIA, portent sur les HMMs [Fox et al., 2006] et les réseaux dynamiques Bayésiens [Infantes et al., 2006]. Nous espérons montrer que notre formulation du problème, par la représentation adoptée pour les gestes, facilite la mise en œuvre du système de reconnaissance final sans dégrader ses performances.

Conclusion

Dans cette thèse, nous avons présenté des travaux sur la détection, le suivi de personnes et la reconnaissance de gestes élémentaires à partir du flux vidéo d'une caméra couleur embarquée sur un robot mobile évoluant en milieu intérieur. Les approches proposées ici, se limitent à une analyse spatio-temporelle dans le plan image afin de répondre aux contraintes propres à notre contexte robotique. L'objectif est de permettre une interaction passive ou active entre une plateforme mobile et les usagers partageant l'environnement.

Dans un chapitre introductif, nous dressons un panorama sommaire des travaux dans le domaine du suivi visuel. Parmi les nombreuses approches de suivi visuel d'humains proposées dans la communauté, nous nous sommes focalisés sur les méthodes de suivi par filtrage particulaire qui sont très adaptées à notre contexte robotique. En effet, les scènes rencontrées sont *a priori* encombrées, dynamiques tandis que les conditions de prise de vue sont très variables. Le filtrage particulaire offre une grande généricité et permet de combiner/fusionner aisément différents sources de mesures. Cependant, il nous semble que la plupart des travaux proposés dans la littérature se limitent à un nombre restreint de primitives visuelles. De plus, peu d'évaluations comparatives entre les différentes stratégies de filtrage particulaire sont proposés. Notre contribution porte sur ces deux derniers points. Les stratégies de fusion de données visuelles et de filtrage associées seront évaluées dans le contexte robotique décrit précédemment.

Le deuxième chapitre est consacré aux méthodes de filtrage particulaire. Nous rappelons tout d'abord quelques généralités sur le filtrage particulaire mettant en œuvre les méthodes de Monte Carlo et présentons l'algorithme générique de filtrage. La nature récursive de cet algorithme peut conduire à une dégénérescence du nuage de particules, induit par l'augmentation dans le temps de la variance inconditionnelle de ses poids. Une étape de rééchantillonnage permet de limiter ce phénomène cependant, l'efficacité de l'algorithme dépend aussi du choix de la fonction d'importance. La fonction d'importance dite optimale positionne les particules en tenant compte de la dynamique du système et de l'observation à l'instant courant et ainsi minimise la variance des poids d'importances mais en pratique, cette fonction n'est généralement pas utilisable. Nous présentons alors différentes fonctions d'importances conduisant à des algorithmes de filtrage aux performences et caractéristiques variables. D'autres méthodes de réduction de la variance des poids d'importance sont aussi abordées et des stratégies basées sur des fonctions d'importances plus complexes afin d'approcher le cas optimal de filtrage sont présentées.

En suivi visuel, les mesures jouent un rôle essentiel dans le fonctionnement du filtre, d'une part dans la définition d'une fonction de vraisemblance des particules, et d'autre part dans la définition d'une fonction d'importance qui détermine la stratégie d'exploration de l'espace d'état. Nous avons donc proposé un ensemble de mesures visuelles et stratégies associées afin de les combiner/fusionner dans un algorithme de filtrage particulaire. Une évaluation de ces stratégies en termes de pouvoir discriminant et temps de calcul indépendament du filtrage a également été proposée en fin de chapitre. Nous avons ainsi mis en évidence que la fonction de mesure fusionnant la mesure de forme à la mesure de distribution de couleur offre un bon compromis vis-à-vis de l'ensemble des critères évalués. Concernant les fonctions d'importance, l'association de détecteurs de visage et de « blobs peau » nous semble pertinente car elle limite le nombre de faux négatifs détectés.

Dans le chapitre 4, nous avons défini trois modalités d'interaction pour un robot censé guider les visteurs dans un musée à savoir : (i) l'interaction proximale, (ii) la mission de guidage et (iii) la surveillance pour interpeller à distance les visiteurs. Nous avons alors décliné différentes stratégies de fusion/combinaison de mesures et de filtrage pour chacune de ces modalités. Celles-ci ont été évaluées sur des séquences représentatives de chacun des scénarios. Les stratégies retenues doivent être robustes aux changements d'apparence, sauts de dynamiques, occultations sporadiques de la cible qui sont fréquents dans notre contexte. Ces évaluations aboutissent à la sélection de stratégies différentes pour chaque modalité. Peu de travaux, à notre connaissance, proposent ce genre d'étude... certes dans un cadre applicatif précis.

Le chapitre 5 traite enfin de l'interaction gestuelle Homme-Robot dans notre contexte. Une reconnaissance de gestes de la main est mise en place pour permettre aux visiteurs de "dialoguer" avec le robot par des gestes. D'après les modalités de suivi définies précédemment, deux modalités d'interaction sont proposées. Une première concerne la demande de changement de but au cours d'une mission. Le visiteur signale au robot son désir de changer d'objectif de visite par un enchaînement de configuration de la main. La deuxième modalité concerne l'apprentissage du prénom du visiteur par le robot durant l'interaction proximale. Dans cette modalité, le visiteur épelle son prénom en réalisant des gestes alphabétiques de la main. Chaque geste effectué correspond à un séquencement naturel de trajectoires canoniques représentant une lettre. Des algorithmes de filtrage particulaire pour les systèmes à sauts Markoviens sont présentés dans ce chapitre. Ils permettent de gérer un indice discret indexant la configuration de la main et/ou un modèle de dynamique canonique dans le vecteur d'état du système. L'évaluation dans ce chapitre porte sur un seul algorithme de filtrage pour lequel on mesure les taux de reconnaissance. Certaines évaluations, realtives à JMSPF restent à effectuer et seront proposées ultérieurement.

Nous avons pu définir dans cette thèse des modalités d'interactions dans le contexte du robot guide de musée. Pour chaque modalité diverses stratégies de suivi ont été envisagées et évaluées. Dans la problématique plus générale du robot personnel, les perspectives et évolutions de ces travaux sont nombreuses.

De nouvelles fonctions de mesure et d'importance peuvent être définies pour améliorer la robustesse du suivi.

Parmi les stratégies de filtrage envisagées, d'autres stratégies peuvent être proposées et évaluées. La stratégie Auxiliary Unscented présentée dans le chapitre 2 par exemple, permet de s'approcher du cas optimal de filtrage. Il est par conséquent indispensable de l'envisager dans notre contexte de suivi de personne afin de l'évaluer et de la comparer aux autres stratégies.

Le suivi visuel en environnement encombré et variable peut conduire à un décrochage du filtre. Une difficulté majeure est de se rendre compte que la cible est perdue. Dans ce contexte, il serait donc intéressant de développer des techniques qui permettent de détecter le décrochage du suivi à partir de l'analyse du comportement du filtre.

Une autre piste de travaux liée à l'environnement et à ces variations est de ne plus considérer la fusion des attributs visuels au même niveau mais de pondérer les mesures selon le contexte. On comprend bien que selon les conditions, certains attributs visuels sont plus pertinents que d'autres. Dans le cas d'un environnement sous éclairé par exemple, l'attribut couleur semble moins approprié que l'attribut forme, il serait donc intéressant de pondérer les mesures de manière adaptée. Il existe déja des travaux dans le domaine [Vermaak et al., 2002b], il semble donc indispensable de prendre en compte ce type de fusion de données et de les évaluer.

Le suivi multi-cible est une piste qui n'a pas encore été explorée dans nos travaux et semble incontournable dans notre contexte, notament dans la perspective de suivre plusieurs personnes de l'environnement ou par exemple pour permettre une interaction gestuelle bimanuelle.

Enfin, nos travaux peuvent se poursuivre vers des approches de suivi 3D par apparence. La thèse de Paulo Menezes actuellement en cours [Menezes et al., 2005] se situe dans cette problématique. Malgré une complexité importante qui demande d'envisager de nouvelles stratégies de filtrage afin de gérer de nombreux degrés de liberté, il nous semble que les techniques de fusion de données présentées dans cette thèse peuvent être utilisées et étendues pour ces approches.

Bibliographie

[Akashi et al., 1977] H. Akashi et H. Kumamoto (1977). Random sampling approach to state estimation in switching environments. *Automatica*, 13 :429–434.

[Albiol et al., 2001] A. Albiol, L. Torres, et E. Delp (2001). An unsupervised color image segmentation algorithm for face detection applications. Dans *Int. Conf. On Image Processing (ICIP'01)*, pages 7–10.

[Alspach et al., 1972] D. L. Alspach et H. W. Sorenson (1972). Nonlinear bayesian estimation using gaussian sum approximation. *IEEE Trans. Automat. Contr.*, pages 439–448.

[Andrieu et al., 2001] C. Andrieu, M. Davy, et A. Doucet (2001). Improved auxiliary particle filtering : applications to time-varying spectral analysis. Dans *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*, pages 309–312.

[Andrieu et al., 2003] C. Andrieu, M. Davy, et A. Doucet (2003). Efficient particle filtering for jump markov systems. application to time-varying autoregressions. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 51 :1762–1770.

[Arulampalam et al., 2002] S. Arulampalam, S. Maskell, N. Gordon, et T. Clapp (2002). A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Trans. On Signal Processing*, 50(2) :174–188.

[Bailly et al., 2005] G. Bailly, L. Br thes, R. Chatila, A. Clodic, J. Crowley, P. Dan s, F. Elisei, S. Fleury, M. Herrb, F. Lerasle, P. Menezes, et R. Alami (2005). Hr+ : Towards an interactive autonomous robot. Dans *Journ es ROBEA*, pages 39–45, Montpellier.

[Batista et al., 1998] J. Batista, P. Peixoto, et A. Helder (1998). Real-time active vision surveillance by integrating peripheral motion detection and foveated tracking. Dans *IEEE Workshop on Visual Surveillance*, Bombay.

[Beucher et al., 1993] S. Beucher et F. Meyer (1993). *Mathematical Morphology in Image Processing, Chapter 12.* Marcel Dekker Inc.

[Birchfield, 1998] S. Birchfield (1998). Elliptical head tracking using intensity gradients and color histograms. Dans *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'98)*, pages 232–237, Santa Barbara.

[Black et al., 1998] M. Black et A. Jepson (1998). Recognizing temporal trajectories using the condensation algorithm. Dans *Int. Conf. On Automatic Face and Gesture Recognition (FGR'98)*, pages 16–21.

[Blake et al., 1993] A. Blake, R. Curwen, et A. Zisserman (1993). A framework for spatio-temporal control in the tracking of visual contours. *Int. Journal of Computer Vision (IJCV'93)*, 11(2) :1265–1278.

[Blake et al., 1998a] A. Blake et M. Isard (1998a). *Active Contours*. Springer.

[Blake et al., 1998b] A. Blake et M. Isard (1998b). *Active Contours*. Springer.

[Boehm et al., 1994] K. Boehm, W. Broll, et M. Sokolewicz (1994). Dynamic gesture recognition using neural networks. Dans *SPIES Conf. on Electronic Imaging Science and Technology*, San Jose (USA).

[Bornstein et al., 1989] H. Bornstein et K. Saulnier (1989). *The Signed English Starter*. CLERC BOOKS. Gallaudet University Press, Washington.

[Bradski, 1998] G. Bradski (1998). Computer vision face tracking as a component of a perceptual user interface. Dans *Workshop on Applications of Computer Vision*, pages 214–219, Princeton.

[Bretzner et al., 2002] L. Bretzner, I. Laptev, et T. Lindeberg (2002). Hand gesture using multi-scale colour features, hierarchical models and particle filtering. Dans *Int. Conf. On Automatic Face and Gesture Recognition(FGR'02)*, pages 405–410, Washington.

[Brèthes et al., 2006a] L. Brèthes, P. Danès, et F. Lerasle (2006a). Particle filtering strategies for visual tracking dedicated to h/r interaction. Dans *Soumis dans : IEEE International Conference on Robotics and Automation (ICRA'06)*.

[Brèthes et al., 2006b] L. Brèthes, P. Danès, et F. Lerasle (2006b). Stratégies de filtrage particulaire pour le suivi visuel de personnes : description et évaluation. Dans *Soumis dans : Reconnaissance Des Formes Et Intelligence Artificielle (RFIA'06)*.

[Brèthes et al., 2005] L. Brèthes, F. Lerasle, et P. Danès (2005). Data fusion for visual tracking dedicated to human-robot interaction. Dans *IEEE Conf. On Robotics and Automation (ICRA'05)*, pages 2087–2092.

[Brèthes et al., 2004a] L. Brèthes, P. Menezes, F. Lerasle, et M. Briot (2004a). Face tracking and hand gesture recognition for human-robot interaction. Dans *Int. Conf. On Robotics and Automation*, pages 1901–1906.

[Brèthes et al., 2004b] L. Brèthes, P. Menezes, F. Lerasle, et M. Briot (2004b). Segmentation couleur et condensation pour le suivi et la reconnaissance de gestes humains. Dans *Reconnaissance Des Formes Et Intelligence Artificielle*, volume 2, pages 967–975.

[Cadoz, 1994] C. Cadoz (1994). Le geste canal de communication homme/machine. la communication « instrumentale ». *Technique et science informatique*, 13(1) :31–61.

[Campillo, 2005] F. Campillo (2005). *Filtrage Linéaire, Non Linéaire et Approximation Particulaire pour le Praticien*. INRIA.

[Canny, 1986] J. Canny (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6) :679–698.

[Carpenter et al., 1999] J. Carpenter, P. Cli, et P. Fearnhead (1999). An improved particle lter for non-linear problems. *iee Proceedings - F : Radar, Sonar and Navigation*, 146 :2–7.

[Casella et al., 1996] G. Casella et C. Robert (1996). Rao-blackwellisation of sampling schemes. *Biometrika*, pages 81–94.

[Chateau et al., 2004] T. Chateau, F. Jurie, R. Marc, et M. Dhome (2004). Reconnaissance de gestes par vision monoculaire temps réel : application à la formation des charges de manoeuvres pour la conduite des ponts polaires. Dans *Reconnaissance Des Formes Et Intelligence Artificielle (RFIA'04)*, volume 2, pages 947–955.

[Chen et al., 2001] H. Chen et T. Liu (2001). Trust-region methods for real-time tracking. Dans *Int. Conf. on Computer Vision (ICCV'01)*, volume 2, pages 717–722, Vancouver.

[Comaniciu et al., 2000] D. Comaniciu, V. Ramesh, et P. Meer (2000). Real-time tracking of non-rigid objects using mean shift. Dans *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00)*, volume 2, pages 142–149, Hilton Head Island, South Carolina.

[Comaniciu et al., 2003] D. Comaniciu, V. Ramesh, et P. Meer (2003). Kernel-based object tracking. Dans *IEEE Trans. Pattern Analysis Machine Intell, (PAMI'03)*, volume 25, pages 564–575.

[Crisan et al., 2002] D. Crisan et A. Doucet (2002). A survey of convergence results on particle filtering for practitioners. *IEEE Trans. Signal Processing*, 50(3) :736–746.

[Cui et al., 1995] Y. Cui, S. D., et J. Weng (1995). Learning-basedhand sign recognition using SHOSLIF-M. Dans *Int. Conf. On Face and Gesture Recognition (FGR'95)*, pages 201–206.

[Davis et al., 1994] J. Davis et M. Shah (1994). Visual gesture recognition. *Vision, Image and Signal Processing*, 141(2) :101–106.

[Davis et al., 2000] L. Davis, V. Philomin, et R. Duraiswani (2000). Tracking humans from a moving platform. Dans *Int. Conf. On Pattern Recognition (ICPR'00)*, volume 4, pages 171–177, Barcelona.

[Doucet, 1998] A. Doucet (1998). On sequential simulation-based methods for bayesian filtering. Technical report, Cambridge University Department of Engineering.

[Doucet et al., 2001a] A. Doucet, N. De Freitas, et N. J. Gordon (2001a). *Sequential Monte Carlo Methods in Practice*. Series Statistics For Engineering and Information Science. Springer-Verlag, New York.

[Doucet et al., 2000] A. Doucet, S. J. Godsill, et C. Andrieu (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3) :197–208.

[Doucet et al., 2001b] A. Doucet, N. Gordon, et V. Krishnamurthy (2001b). Particle filters for state estimation oj jump markov linear systems. *IEEE Transaction on signal processing*, 49(3) :613–624.

[Fox et al., 2006] M. Fox, M. Ghallab, G. Infantes, et D. Long (2006). Robot introspection through learned hidden markov models. *Artificial Int. Journal (AIJ'06), à paraître.*

[Freund et al., 1995] Y. Freund et R. Schapire (1995). A decision-theoretic generalization of on-line learning and an application to boosting. Dans *European Conference on Computational Learning Theory*, pages 23–37.

[Gavrila, 1998] D. M. Gavrila (1998). Multi-feature hierarchical template matching using distance transforms. Dans *Int. Conf. On Pattern Recognition (ICPR'98)*, pages 439–444.

[Gavrila, 1999] D. M. Gavrila (1999). The visual analysis of human movement : A survey. *Computer Vision and Image Understanding (CVIU'99)*, (1) :82–98.

[Gavrila, 2000] D. M. Gavrila (2000). Pedestrian detection from a moving vehicle. Dans *European Conf. On Computer Vision (ECCV'00)*, Dublin.

[Giebel et al., 2004] J. Giebel, D. M. Gavrila, et C. Schnorr (2004). A bayesian framework for multi-cue 3D object. Dans *Europ. Conf. on Computer Vision (ECCV'04)*, Pragues.

[Gordon et al., 1993] N. Gordon, D. Salmond, et A. Smith (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2) :107–113.

[Gustafsson et al., 2002] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Janson, R. Karlsson, et P.-J. Nordlund (2002). Particle filters for positioning, navigation and tracking. *IEEE Transactions on Signal Processing*, 50(2) :425–437.

[Hammersley et al., 1954] J. Hammersley et K. Morton (1954). Poor man's monte carlo. *Journal of the Royal Statistical Society B*, 16 :23–38.

[Handschin et al., 1970] J. Handschin et D. Mayne (1970). Monte carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International Journal of Control*, 9(5) :547–559.

[Haritaogly et al., 2000] I. Haritaogly, D. Harwood, et L. Davis (2000). W4 : Real-time surveillance of people and their activities. *IEEE Trans. On Pattern Analysis Machine Intelligence*, 8(22) :809–830.

[Herodotou et al., 1998] N. Herodotou, K. Plataniotis, et A. Venetsanopoulos (1998). A color segmentation scheme for object-based video coding. Dans *IEEE Symp. on Advances in Signal Filtering and Signal processing*, pages 25–30.

[Horn et al., 81] B. K. P. Horn et B. G. Schunck (81). Determining optical flow. *Artificial Intelligence*, 16 :185–203.

[Hue et al., 2000] C. Hue, J. Le Cadre, et P. Perez (2000). Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3) :791–812.

[Huttenlocher et al., 1993] D. P. Huttenlocher, J. J. Noh, et W. J. Rucklidge (1993). Tracking non-rigid objects in complex scenes. Dans *Int. Conf. On Computer Vision (ICCV'93)*, volume 1, pages 93–101, Berlin.

[Infantes et al., 2006] G. Infantes, F. Ingrand, et M. Ghallab (2006). Apprentissage de modèle d'activité stochastique pour la planification et le contrôle d'exécution. Dans *Reconnaissance des Formes et Intelligence Artificielle (RFIA'06), à paraître.*

[Isard et al., 1996a] M. Isard et A. Blake (1996a). Contour tracking by stochastic propagation of conditional density. Dans *European Conf. On Computer Vision*, Cambridge.

[Isard et al., 1996b] M. Isard et A. Blake (1996b). Learning dynamics of complex motions from image sequences. Dans *European Conf. on Computer Vision (ECCV'96)*, pages 357–368, Cambridge.

[Isard et al., 1996c] M. Isard et A. Blake (1996c). Visual tracking by stochastic propagation of conditional density. Dans *European Conf. On Computer Vision*, pages 343–356, Cambridge.

[Isard et al., 1998a] M. Isard et A. Blake (1998a). Condensation – conditional density propagation for visual tracking. *Int. J. Comput. Vision*, 29(1) :5–28.

[Isard et al., 1998b] M. Isard et A. Blake (1998b). Icondensation : Unifying low-level and high-level tracking in a stochastic framework. Dans *ECCV '98 : Proceedings of the 5th European Conference On Computer Vision-Volume I*, pages 893–908, London, UK. Springer-Verlag.

[Isard et al., 1998c] M. Isard et A. Blake (1998c). A mixed-state condensation tracker with automatic model-switching. Dans *ICCV '98 : Proceedings of the Sixth International Conference On Computer Vision*, page 107, Washington, DC, USA. IEEE Computer Society.

[Isard et al., 2001] M. Isard et J. MacCormick (2001). Bramble : A bayesian multiple-blob tracker. Dans *Proc. Int. Conf. Computer Vision*, pages 34–41.

[Isard et al., 1998d] M. A. Isard et A. Blake (1998d). A mixed-state condensation tracker with automatic model-switching. Dans *Int. Conf. On Computer Vision*, pages 107–112, Bombay.

[Jepson et al., 2001] A. Jepson, D. Fleet, et T. El-Maraghi (2001). Robust online appearance models for visual tracking. Dans *Int. Conf. On Computer Vision and Pattern Recognition (CVPR'01)*, pages 415–422.

[Jones et al., 1998] M. Jones et J. Rehg (1998). Color detection. Rapport technique, Compaq Cambridge Research Lab.

[Jones et al., 1999] M. J. Jones et J. M. Rehg (1999). Statistical color models with application to skin detection. Dans *Int. Conf. On Computer Vision and Pattern Recognition (CVPR'99)*, pages 274–280.

[Julier, 2002] S. Julier (2002). The scaled unscented transformation. Dans *American Control Conference (ACC'02).*

[Julier et al., 1996] S. Julier et J. Uhlmann (1996). A general method for approximating nonlinear transformations of probability distributions. Rapport technique, RRG, Dept. of Engineering Science, University of Oxford.

[Julier et al., 1997] S. Julier et J. Uhlmann (1997). A new extension of the kalman filter to nonlinear systems. Dans *In Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, Orlando.

[Jurie et al., 2002] F. Jurie et M. Dhome (2002). Hyperplane approximation for template matching. *IEEE Trans. On Pattern Analysis Machine Intelligence (PAMI'02)*, 24(7) :996–1000.

[Kapuscinski et al., 2001] T. Kapuscinski et M. Wysocki (2001). Hand gesture recognition for man-machine interaction. Dans *Robot Motion and Control*, pages 91–96.

[Kawato et al., 2000] S. Kawato et J. Ohya (2000). Automatic skin-color distribution extraction for face detection and tracking. Dans *Int. Conf. on Signal Processing (ICSP'00)*, volume 2, pages 1415–1418.

[Kervrann et al., 1996] C. Kervrann et F. Heitz (1996). Apprentissage non supervisé et suivi de modèles déformables dans une séquence d'images. Dans *Reconnaissance Des Formes Et Intelligence Artificielle (RFIA'96)*, Rennes.

[Kitagawa, 1996] G. Kitagawa (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1) :1–25.

[Kong et al., 1994] A. Kong, J. Liu, et W. Wong (1994). Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association*, 89(425) :278–288.

[Konrad, 2000] J. Konrad (2000). *Motion Detection and Estimation.* Handbook of Image and Video Processing.

[Kotecha et al., 2003] J. Kotecha et P. Djuric (2003). Gaussian sum particle filtering. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on*, 51(10) :2602 – 2612.

[Kwok et al., 2004] C. Kwok, D. Fox, et M. Meil (2004). Real-time particle filter. *Proceedings of the ieee (issue on State Estimation)*, 92(2).

[Lee et al., 2002] J. Lee et S. Yoo (2002). An elliptical boundary model for skin color detection. Dans *Int. Conf. on Imaging Science, Systems and Technology (CISST'02)*, pages 100–106, Las Vegas.

[Li et al., 2002] P. Li et T. Zhang (2002). Visual contour based on sequential importance sampling/resampling algorithm. Dans *Int. Conf. On Pattern Recognition (ICPR'02)*, pages 564–568.

[Liang et al., 1995] R. Liang et M. Ouhhyoung (1995). A real-time continuous alphabetic sign language to speech conversion VR system. *Computer Graphics Forum*, 14(3) :67–76.

[Lichtenauer et al., 2004] J. Lichtenauer, M. J. T. Reinders, et E. A. Hendriks (2004). Influence of the observation likelihood function on particle filtering performance in

tracking applications. Dans *Automatic Face and Gesture Recognition (FGR'04)*, pages 767–772.

[Lienhart et al., 2002] R. Lienhart et J. Maydt (2002). An extended set of haar-like features for rapid object detection. Dans *Int. Conf. On Image Processing (ICIP'02)*, pages 900–903, Thessaloniki.

[Lindeberg, 1998] T. Lindeberg (1998). Feature detection with automatic scale selection. *Int. Journal of Computer Vision*, 30(2) :77–116.

[Liu et al., 2004] Y. Liu et Y. Jia (2004). A robust hand tracking for gesture-based interaction of wearable computers. Dans *Int. Symp. on Wearable Computers (ISWC'04)*.

[MacCormick, 2000] J. MacCormick (2000). *Probabilistic modelling and stochastic algorithms for visual localisation and tracking*. Thèse de doctorat, Department of Engineering Science, University of Oxford.

[MacCormick et al., 1999] J. MacCormick et A. Blake (1999). A probabilistic exclusion principle for tracking multiple objects. Dans *Proc. Int. Conf. Computer Vision (ICCV'99)*, pages 572–578.

[MacCormick et al., 2000] J. MacCormick et M. Isard (2000). Partitioned sampling, articulated objects, and interface-quality hand tracking. Dans *ECCV '00 : Proceedings of the 6th European Conference on Computer Vision-Part II*, pages 3–19, London, UK. Springer-Verlag.

[Marcel et al., 2000] S. Marcel, O. Bernier, J. Viallet, et D. Collobert (2000). Hand gesture recognition using input/output hidden markov models. Dans *Int. Conf. on Face and Gestures Recognition (FGR'00)*, pages 456–461.

[Menezes et al., 2004] P. Menezes, J. Barreto, et J. Dias (2004). Face tracking based on haar-like features and eigenfaces. Dans *Int. Conf. on Robotics and Automation (ICRA'04)*, pages 1888–1893.

[Menezes et al., 2003] P. Menezes, L. Brèthes, F. Lerasle, P. Danès, et J. Dias. (2003). Visual tracking of silhouettes for human-robot interaction. Dans *Int. Conf. On Advanced Robotics (ICAR'01)*, volume 2, pages 971–976.

[Menezes et al., 2005] P. Menezes, F. Lerasle, J. Dias, et R. Chatila (2005). Suivi visuel de structures articul es 3d par filtrage particulaire. Dans *ORASIS*.

[Merwe et al., 2000] R. v. d. Merwe, A. Doucet, N. d. Freitas, et E. Wan (2000). The unscented particle filter. Rapport technique, Rapport technique cued/f-infeng/tr380 University of Cambridge, Enguneering department.

[Merwe et al., 2001] R. v. d. Merwe et E. Wan (2001). The square-root unscented kalman filter for state and parameter-estimation. Dans *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, Utah.

[Metropolis et al., 1953] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, et E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21 :1087–1092.

[Metropolis et al., 1949] N. Metropolis et S. Ulam (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(335–341).

[Millet, 2005] A. Millet (2005). *Méthodes de Monte Carlo*. Université Paris 6.

[Moeslund et al., 2001] T. Moeslund et E. Granum (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding (CVIU'01)*, 81(3).

[Moulines, 2002] E. Moulines (2002). *Hidden Markov Models and Particle Filters*.

[Nguyen et al., 2001] H. Nguyen et A. Smeulders (2001). Occlusion robust adaptative template matching. Dans *Int. Conf. on Computer Vision (ICCV'01)*, volume 1, pages 678–683.

[Nummiaro et al., 2002] K. Nummiaro, E. Koller-Meier, et L. V. Gool (2002). Object tracking with an adaptative color-based particle filter. Dans *Symp. For Pattern Recognition of the DAGM*, pages 353–360.

[Nummiaro et al., 2003] K. Nummiaro, E. Koller-Meier, et L. V. Gool (2003). An adaptative color-based particle filter. *Journal of Image and Vision Computing*, 21 :90–110.

[Ohta et al., 1980] Y. Ohta, T. Kanade, et T. Sakai (1980). Color information for region segmentation. *Computer Graphics and Image Processin (CGIP'80)*, 10(13) :222–241.

[Papanikolopoulos et al., 1993] N. Papanikolopoulos, P. Khosla, et T. Kanade (1993). Visual tracking of a moving target by a camera mounted on a robot. *IEEE Trans. on Robotics and Automation (TRA'93)*, 9 :14–35.

[Paragios et al., 2000] N. Paragios et R. Deriche (2000). Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. Pattern Analysis Machine Intelligence(PAMI'00)*, 22(3) :850–863.

[Pavlovic et al., 2000] V. Pavlovic, J. Rehg, et J. MacCormick (2000). Impact of dynamic model learning on classification of human motion. Dans *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, Hilton Head Island.

[Pavlovic et al., 1997] V. Pavlovic, R. Sharma, et T. S. Huang (1997). Visual interpretation of hand gestures for human-computer interaction : A review. *IEEE Trans. On Pattern Analysisand Machine Intelligence (PAMI'97)*, 19(7) :677–695.

[Phung et al., 2005] S. L. Phung, S. A.. Bouzerdoum, et S. D.. Chai (2005). Skin segmentation using color pixel classification : Analysis and comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, 27 :148–154.

[Pitt et al., 2001] M. Pitt et M. Shephard (2001). *Auxiliary variable based particle filters*, chapitre 13. Springer-Verlag.

[Pitt et al., 1999] M. K. Pitt et N. Shephard (1999). Filtering via simulation : Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446) :590–599.

[Plataniotis et al., 2000] K. Plataniotis et A. Venetsanopoulos (2000). *Color Image Processing and Applications, Chapter 4*. Springer.

[Pérez et al., 2002] P. Pérez, C. Hue, J. Vermaak, et M. Gangnet (2002). Color-based probabilistic tracking. Dans *Eur. Conf. on Computer Vision, ECCV'2002, LNCS 2350*, pages 661–675, Copenhaguen, Denmark.

[Pérez et al., 2004] P. Pérez, J. Vermaak, et A. Blake (2004). Data fusion for visual tracking with particles. *Proc. IEEE*, 92(3) :495–513.

[Quek, 1994] F. K. H. Quek (1994). « Toward a Vision-Based Hand Gesture Interface ». *Virtual Reality Software and Technology*.

[Rabiner, 1989] L. Rabiner (1989). A tutorial on hmm and selected applications in speech recognition. *In Proc. IEEE*, 77(2) :257–286.

[Rittscher et al., 1999] J. Rittscher et A. Blake (1999). Classification of human body motion. Dans *Int. Conf. On Computer Vision (ICCV'99)*, Kerkyra (Greece).

[Rui et al., 2001] Y. Rui et Y. Chen (2001). Better proposal distributions : Object tracking using unscented particle filter. Dans *Int. Conf. On Computer Vision and Pattern Recognition (CVPR'01)*, pages 786–793.

[Schwerdt et al., 2000] K. Schwerdt et J. L. Crowley (2000). Robust face tracking using color. Dans *Int. Conf. On Face and Gesture Recognition (FGR'00)*, pages 90–95, Grenoble, France.

[Shan et al., 2004] C. Shan, Y. Wei, X. Qiu, et T. Tan (2004). Gesture recognition using temporal template based trajectories. Dans *Int. Conf. on Pattern Recognition (ICPR'04)*.

[Sminchisescu, 2002] C. Sminchisescu (2002). *Estimation Algorithms for Ambiguous Visual Models*. Thèse de doctorat, Institut National Polytechnique de Grenoble.

[Spengler et al., 2001] M. Spengler et B. Schiele (2001). Towards robust multi-cues integration for visual tracking. Dans *Workshop On Computer Vision Systems*.

[Stenger et al., 2003] B. Stenger, A. Thayananthan, P. Torr, et R. Cipolla (2003). Filtering using a Tree-based Estimator. Dans *Int. Conf. on Computer Vision (ICCV'03)*, pages 1063–1070.

[Terrillon et al., 2000] J. Terrillon, M. Shirazi, H. Fukamachi, et S. Akamatsu (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. Dans *Int. Conf. on Face and Gesture Recognition (FGR'00)*, pages 54–61.

[Thayananthan et al., 2003] A. Thayananthan, B. Stenger, P. Torr, et R. Cipolla (2003). Learning a kinematic prior for tree-based filtering. Dans *British Machine Vision Conf. (BMVC'03)*, volume 2, pages 589–598, Norwick.

[Thiel, 1994] E. Thiel (1994). *Les Distances de Chanfrein en Analyse des images : Fondements et Applications*. Thèse de doctorat, Universit Joseph Fourier, Grenoble I.

[Torma et al., 2003] P. Torma et C. Szepesvári (2003). Sequential importance sampling for visual tracking reconsidered. Dans *AI and Statistics*, pages 198–205.

[Triesch et al., 2002] J. Triesch et der V. Malsburg (2002). Classification of hand postures against complex backgrounds using elastic graph matching. *Image and Vision Computing*, (20) :937–943.

[van der Merwe et al., 2003] van der R. Merwe et E. Wan (2003). Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. Dans *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong. IEEE.

[Vermaak et al., 2002a] J. Vermaak, C. Andrieu, A. Doucet, et S. Godsill (2002a). Particle methods for bayesian modeling and enhancement of speech signals. *IEEE Transactions on Speech and Audio Processing*, 10(3) :173–185.

[Vermaak et al., 2002b] J. Vermaak, P. P&#233 ;rez, M. Gangnet, et A. Blake (2002b). Towards improved observation models for visual tracking : Selective adaptation. Dans *Proceedings of the 7th European Conference on Computer Vision-Part I (ECCV '02)*, pages 645–660, London, UK. Springer-Verlag.

[Vezhnevets et al., 2003] V. Vezhnevets, V. Sazonov, et A. Andreeva (2003). A survey on pixel-based skin color detection techniques. *Proc. Graphicon-2003*, pages 85–92.

[Viola et al., 2001] P. Viola et M. Jones (2001). Rapid object detection using a boosted cascade of simple features. Dans *Int. Conf. On Computer Vision and Pattern Recognition*.

[Wagener et al., 2003] D. Wagener et B. Herbst (2003). Face tracking for expressions simulations. Dans *Int. Conf. On Computer Systems and Technologies (ICCST'03)*, pages 253–259.

[Wu et al., 1999] Y. Wu et T. Huang (1999). Vision-based gesture recognition : a review. Dans *Int. Gesture Workshop on Gesture-based Communication in Human-Computer Interaction*, pages 103–115.

[Wu et al., 2001] Y. Wu et T. Huang (2001). A co-inference approach to robust visual tracking. Dans *Int. Conf. on Computer Vision (ICCV'01)*, volume 2, pages 26–33.

[Yang et al., 1997] J. Yang, Y. Xu, et C. Chen (1997). Human action learning via hidden markov model. *IEEE Trans. Systems, Man, Cybernetics*, 27(1) :34–44.

[Yang et al., 2002] M. Yang, D. Kriegman, et N. Ahuja (2002). Detecting faces in images : A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1) :34–58.

[Yoon et al., 2001] H. Yoon, J. Soh, Y. Bae, et H. Yang (2001). Hand gesture recognition using combined features of location, angle and velocity. *Pattern Recognition (PR'01)*, (34) :1491–1501.

[Zarit et al., 1999] B. D. Zarit, B. J. Super, et K. H. Quek (1999). Comparison of five color models in skin pixel classification. Dans *ICCV'99 International Workshop On Recognition, Analysing, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'99)*, pages 58–63.

[Zenzo, 1986] S. D. Zenzo (1986). A note on the gradient of a multi-image. *Int. Journal of Computer Graphics and Image Processing (CVGIP'86)*, 33 :116–125.

Table des figures

Suivi visuel par filtrage particulaire. Application à l'interaction Homme-Robot.

Résumé : Un défi majeur de la Robotique aujourd'hui est sans doute celui du robot personnel, capable de rendre service à l'Homme. De nombreux travaux de recherche dans le domaine sont axés sur le développement de robots autonomes destinés à évoluer dans des environnements de grandes dimensions en présence de public. Cette perspective pose naturellement le problème de l'interaction et de la relation entre l'Homme et le robot. En effet, lors de sa navigation, le robot doit être capable de détecter et de prendre en compte de manière explicite la présence de personnes dans son voisinage pour les éviter ou leur céder le passage, le but étant de faciliter et de sécuriser leur déplacements. De plus, il doit disposer de capacités d'interaction telles que la reconnaissance de gestes permettant à l'Homme de communiquer avec lui. Cette thèse porte plus spécifiquement sur la détection et le suivi de personnes ainsi que la reconnaissance de gestes élémentaires à partir du flot vidéo d'une caméra couleur embarquée sur le robot.

Le filtrage particulaire est très adapté à ce contexte. Il permet de s'affranchir de toute hypothèse restrictive quant aux distributions de probabilités entrant en jeu dans la caractérisation du problème. De plus, ce formalisme permet de combiner/fusionner aisément différentes sources de mesures. Malgré ce constat, la fusion de données par filtrage particulaire nous semble assez peu exploitée et souvent confinée à un nombre relativement restreint de primitives visuelles. Nous proposons différents schémas de filtrage, où l'information visuelle est prise en compte dans les fonctions d'importance et de vraisemblance au moyen de primitives forme, couleur et mouvement image. Nous évaluons alors quelles combinaisons de primitives visuelles et d'algorithmes de filtrage répondent au mieux aux modalités d'interaction envisagées pour notre robot "guide de musée", qui est censé interpeller les visiteurs, interagir avec eux et les guider.

Notre dernière contribution porte sur la reconnaissance de gestes symboliques permettant de communiquer avec le robot. Une stratégie de filtrage particulaire efficace est proposée afin de suivre et reconnaître simultanément des configurations de la main et des dynamiques gestuelles dans le flot vidéo.

Mots Clefs : interaction visuelle homme-robot, détection, suivi, filtrage particulaire.

Particle filtering-based visual tracking. Application to Human-Robot interaction.

Abstract : Nowadays, the major challenge of Robotics is certainly the personal robot which is able to be of use to human. Many researches in this field are axed on the development of autonomous robots intended to evolve in large human environments. Obviously, this perspective shows the problem of the interaction and the relation between men and robots. Indeed, during the navigation, the robot must be able to detect humans presence in its vicinity and to take them into account, by explicit manner, in order to avoid them or to let them pass : the goal is to facilitate and to make safe their displacements. Moreover, it must have capacities of interaction such as gestures' recognition which allow Men to communicate with him. This thesis is focused more specifically on the detection and the tracking of people and also on the recognition of elementary gestures from video stream of a color camera embeded on the robot.

Particle filter is well suited to this context. It allows to avoid any restrictive assumption about the probabilities distributions which are throw in the characterization of the problem. Moreover, this formalism enables a straight combination/fusion of several measurement cues. Despite of this observation, particle filter data fusion seems to us to be little exploited and often confined to a relatively restricted number of visual cues. We propose various filtering strategies where visual information such as shape, color and motion are taken into account in the importance function and the measurement model. We compare and evaluate these filtering strategies in order to show which combination of visual cues and particle filter algorithm are more suitable to the interaction modalities that we consider for our tour-robot which is supposed to hail visitors, to interact with them and to guide them.

Our last contribution relates to the recognition of symbolic gestures which enable to communicate with the robot. An efficient particle filter strategy is proposed in order to track the hand and to recognize at the same time its configuration and gesture dynamic in video stream.

Keywords : human-robot visual interaction, detection, tracking, particle filter.