



---

# Implémentation d'un comportement proactif d'une plateforme robotique

---

*Auteur :*

M. Fabrice POIRIER

*Tuteur entreprise :*

J. CHODOROWSKI

*Tuteur école :*

J. ZERR

*Entreprise :*

Orange Labs, Lannion

Version du  
21 août 2019



# Remerciements

Je tiens à remercier Rémi pour m'avoir transmis l'offre de stage, m'avoir soutenu et supporté tout au long du stage. Je remercie Jacques et Benoît pour m'avoir offert ce soutien et cette liberté dans mes actions, me permettant de me focaliser sur le plus important. Je remercie également Thibaut et Matheus pour donner à cette équipe une cohésion et une bonne ambiance unique.

Je remercie enfin l'ensemble de l'équipe HFC qui m'a accueilli chaleureusement et sans condition, ce qui m'a permis de vivre mon séjour estival à Lannion en toute quiétude.



# Résumé

Dans le cadre de la robotique sociale, la qualité des interactions entre la plateforme robotique et l'utilisateur influent grandement sur l'appréciation du service rendu par la machine. Il convient donc de fiabiliser ces interactions sur plusieurs aspects : la prise d'information, l'engagement, et l'interaction en elle-même. L'engagement influence l'interaction, il est donc nécessaire d'étudier les possibilités permettant de le rendre le plus naturel et agréable possible. Pour cela la qualité de la prise d'informations est d'un intérêt crucial pour ne pas faire de mauvaises interprétations basées sur des données faussées ou incomplètes.

Ce rapport propose dans un premier temps une méthode de suivi et de prises d'informations sur des personnes à l'aide d'une caméra de profondeur. Cette méthode fonctionne sur un couplage de trois technologies : la première détecte la présence et la posture de personnes, la deuxième gère la localisation et le suivi de ces personnes tandis que la dernière assure l'identification par reconnaissance faciale des personnes suivies.

Dans un second temps, la mise en œuvre d'un comportement proactif est étudiée en prenant en compte le cadre d'utilisation, et une première approche est implémentée.

Des expérimentations en situation réelle sur un robot humanoïde montrèrent des résultats encourageants quant à la méthodologie employée. Les nombreuses données récoltées constituent une base solide pour poursuivre l'implémentation d'un comportement proactif en adéquation avec l'état de l'art. L'implémentation s'inscrit dans une logique modulaire ce qui permet une évolution aisée vers des services plus complexes, voire l'utilisation sur d'autres systèmes.



# Abstract

In the context of social robotics, the quality of interactions between the robotic platform and the user greatly influences the appreciation of the service provided by the machine. It is therefore necessary to make these interactions more reliable on several levels : information gathering, engagement, and the interaction itself. Engagement influences interaction, it is thus necessary to study the possibilities in order to make it as natural and pleasant as possible. For this reason, the quality of information gathering is of crucial importance in order not to make misinterpretations based on inaccurate or incomplete data.

At first, This report proposes a method for tracking and collecting information about people using a depth camera. This method works on a coupling of three technologies : the first detects the presence and attitude of people, the second handle the localization and tracking of these people while the last one ensures facial recognition identification of followed persons.

In a second step, the implementation of a proactive behaviour is studied taking into account the framework of use, and a first approach is implemented.

Real-life experiments on a humanoid robot showed encouraging results in regards to the methodology used. The extensive data collected provides a solid basis for further implementation of proactive behaviour in line with the state of the art. The implementation is based on a modular logic which allows an easy evolution towards more complex services, or even use on other devices.



# Table des matières

<b>1</b>	<b>Algorithme de suivi</b>	<b>15</b>
1.1	État de l'art . . . . .	15
1.2	Fonctionnement . . . . .	16
1.3	Critère de suivi . . . . .	17
1.3.1	Suivi spatial des « squelettes » . . . . .	17
1.3.2	Suivi temporel des « squelettes » . . . . .	20
<b>2</b>	<b>Algorithme d'identification</b>	<b>21</b>
2.1	Objectifs . . . . .	21
2.2	Fonctionnement . . . . .	21
2.3	Critère de qualité . . . . .	22
2.4	Critère de reconnaissance . . . . .	22
2.4.1	Reconnaissance faciale . . . . .	22
2.4.2	Création de modèle . . . . .	22
2.5	Discussion sur l'approche globale . . . . .	23
<b>3</b>	<b>Algorithme de proactivité</b>	<b>25</b>
3.1	Proactivité . . . . .	25
3.1.1	Formalisation . . . . .	25
3.1.2	Limites . . . . .	26
3.2	État de l'art . . . . .	26
3.3	Implémentation . . . . .	27
3.3.1	Fonctionnement . . . . .	27
3.3.2	Limites . . . . .	28
<b>4</b>	<b>Expérimentations</b>	<b>29</b>
4.1	Cas d'utilisation . . . . .	29
4.2	Résultats . . . . .	30
4.3	Interprétation . . . . .	32
<b>5</b>	<b>Conclusion et perspectives</b>	<b>33</b>
5.1	Conclusion . . . . .	33

5.2	Mémorisation et données personnelles . . . . .	33
5.3	Avenir du projet . . . . .	34
<b>A</b>	<b>Actionlib enhanced</b>	<b>39</b>
<b>B</b>	<b>item-tracking</b>	<b>41</b>
<b>C</b>	<b>Logigramme du cas d'utilisation</b>	<b>43</b>

# Introduction

## Orange et Orange Labs

Orange, anciennement France Telecom, est une entreprise française au rayonnement international. Elle emploie 151 000 salariés pour 264 millions de clients<sup>1</sup>. Son activité, anciennement restreinte à la téléphonie (fixe et mobile) et au réseau (offre internet), s'est ouverte au secteur bancaire (avec Orange Money, Orange Cash et Orange Bank), aux services télévisés (Orange Cinéma Série) et musicaux (partenariat avec Deezer), et l'entreprise s'est dotée d'une présence marine avec des navires câbliers. L'ensemble des secteurs sont traités en trois entités : le service commercial, le service technique et le service R&D.

Orange Labs est la division Recherche d'Orange qui compte plus de 5000 chercheurs, ingénieurs, techniciens et designers. Elle est elle-même séparée en trois entités : Orange Labs Services (OLS) développe des services, Orange Labs Network (OLN) dédiée au réseau de télécommunication, et Orange Labs Research (OLR) s'occupe de la recherche.

## Équipe robotique

L'équipe en charge de la robotique à Orange Labs a pour objectif de faire de la veille technologique dans la robotique sociale et mobile. De ce fait il s'agit de se former au niveau de l'état de l'art du domaine, et de fournir une expertise pour les autres projets confrontés à ces sujets. Pour cela il est nécessaire d'avoir accès à des robots, ainsi que les algorithmes et concepts nécessaires pour les utiliser efficacement. L'équipe étant à effectif réduit il s'agit donc de bien choisir les composantes les plus intéressantes du domaine.

## Robot Operating System (ROS)

ROS est un intergiciel (*middleware*) open-source dédié à la robotique fournissant un formalisme normalisé facilitant l'échange d'informations hétérogènes (capteurs diverses, algorithmie, réseau, contrôle d'actionneur, ...). Il est capable de faire fonctionner chaque brique sur un appareil différent, ce qui permet de déporter les calculs sur un serveur externe lorsque la capacité de calcul des ordinateurs embarqués est insuffisante (ce qui est souvent le cas). Il fournit en plus

---

1. <https://www.orange.com/fr/Groupe/Chiffres-cles/Decouvrez-les-chiffres-clefs-du-Groupe-Orange>

un ensemble d'outils permettant la visualisation du fonctionnement du système. Les langages utilisés sont C++ et Python. Finalement, la communauté derrière ROS est très active, très large et variée (il est utilisé dans les domaines universitaires, entrepreneuriales, ou encore de recherche).

Pour ses caractéristiques, ROS est la plateforme privilégiée par l'équipe.

## Bases roulantes

Il y a plusieurs bases roulantes dans les locaux d'Orange Labs, chacune ayant son intérêt et son cadre d'utilisation.

Turtlebot<sup>2</sup> est une famille de robots roulants en kit, dont les codes source sont open-source. Ils ont été conçus pour fonctionner sous ROS. La prise en main est relativement aisée, mais il s'agit de bases de recherche, ce qui ne convient pas sur le long terme dans un cadre professionnel.

Geko<sup>3</sup> est un autre robot à quatre roues de plus grande taille.

Il est défini par :

- Une bonne finition
- Un interfaçage avec ROS
- Une bonne motricité
- Une caméra mobile embarquée

Ce robot a été choisi comme base roulante par l'équipe pour ces caractéristiques, et pour la bonne relation entretenue avec le constructeur.

## Robots humanoïdes

L'objectif étant de réaliser des services sociaux par le biais de robots, il est tout à fait naturel de travailler avec des robots humanoïdes.

Pepper<sup>4</sup> est un robot humanoïde développé par SoftBank Robotics. Il est utilisé mondialement dans la recherche et dans certains événements. Il s'agit d'un robot industrialisé avec la possibilité de le programmer à volonté. Il est utilisé au sein de l'équipe pour des expériences sur l'acceptation des robots dans la sphère privée.

Waldo<sup>5</sup> est un robot humanoïde développé par Immersive Robotics. Son prix d'origine relativement faible pour ses fonctionnalités fut la raison de son achat. La version de l'équipe fut toutefois fortement modifiée pour augmenter ses capacités. Il embarque actuellement :

- Un ordinateur embarqué NUC sous Linux permettant d'interagir avec le robot sous ROS ;
- Deux caméras dans la tête ;
- Une Kinect pour la vue en profondeur ;
- Un Lidar pour la navigation ;
- Une tablette avec caméra sur le torse ;

---

2. <https://www.turtlebot.com/>

3. <https://www.nuzoo.it/it/prodotti-robot/geko-robot-piattaforma-sviluppo>

4. <https://www.softbankrobotics.com/emea/fr/pepper>

5. <http://immersive-robotics.com/fr/>

- Un bras à 7 degrés de liberté.

Lors du stage la Kinect fut remplacée par la caméra Intel Realsense d435i. Celle-ci cumule de nombreux avantages :

- Elle possède une caméra de profondeur, une caméra couleur, et une IMU ;
- Elle s'alimente et transfère les flux vidéo par un seul port USB C ;
- Elle est paramétrable en fréquence d'images, résolution et en types de données transmis avec une variété de filtres applicables sur l'image de profondeur ;
- Son constructeur maintient à jour un dépôt Git permettant d'utiliser et de paramétrer rapidement la caméra sur ROS (Linux).

Waldo est le robot utilisé dans le cadre de ce stage pour ses capacités de captation, d'interaction et le contrôle total sur son architecture.

## Objectifs

Les plateformes robotiques précédemment citées furent acquises dans le but d'étudier l'intérêt de la robotique dans des services développés pour Orange, voire d'initier la création d'un service.

La direction de la recherche au sein de l'équipe mena à une réflexion sur l'interaction Homme-Machine, et à la méthode pour initier cette interaction de manière naturelle. Il s'agit donc de définir ce qu'est une interaction sociale ainsi que ses déclencheurs. Le cadre de ce stage est d'étudier les méthodes possibles pour commencer automatiquement une interaction de manière proactive (voir partie 3.1) sur une plateforme robotique. Bien que l'interaction en elle-même ne soit pas l'objectif de ce rapport, il s'agit de définir si une interaction est voulue et de l'engager dans le cas d'une personne ne signifiant pas explicitement sa volonté d'interagir.

Dans un premier temps l'objectif fut de mettre en place une brique technologique structurant les données captées par le robot. Dans un second temps, il s'agit d'implémenter un comportement proactif à partir des données recueillies.

## Motivations

En robotique l'adaptation du comportement en fonction de l'environnement est une nécessité, et plus particulièrement en robotique sociale où l'interaction Homme/Machine en est un des thèmes principaux. De nombreuses recherches ont été menées sur ce domaine, à la fois sur l'aspect ergonomique, sociologique, et robotique. Il n'y a toutefois pas de structure logique assez complexe ou généraliste pour englober les objectifs précédemment cités. Il s'agit donc d'apporter une solution technique s'appuyant d'une part sur l'état de l'art sur les modalités d'interactions Homme/Machine, d'autre part sur le traitement de l'information provenant de différentes sources (image couleur, image de profondeur, son, vidéo, LiDAR), tout en restant suffisamment générique pour alimenter de futurs travaux. L'ensemble de ces implémentations

fut donc réalisé de façon à pouvoir facilement interchanger les technologies utilisées avec de potentielles futures meilleures technologies.

## Gestion du projet

La gestion du projet robotique suit une démarche Agile<sup>6</sup>, et plus particulièrement la méthode Kanban. Il s'agit d'une méthodologie permettant une meilleure adaptabilité dans le cadre de développement informatique, et qui s'articule autour de plusieurs concepts. Il s'agit de découper le projet en sous-étapes fonctionnelles mais n'offrant pas forcément l'entièreté de l'application. Chaque sous-étape est séparée en tâches, pouvant être traitées séparément. Chacune de ces tâches passent dans plusieurs états de développement. Dans le cadre du projet ces états furent : « À faire », « en cours » et « fait ».

Concernant l'équipe robotique, les travaux effectués s'inscrivent dans des domaines de compétences présents dans d'autres équipes d'Orange. Ainsi, Des collaborations inter-équipes furent établies pour alimenter les travaux de chacun.

---

6. Plus de détails sur <https://agilemanifesto.org/>

# Chapitre 1

## Algorithme de suivi

### 1.1 État de l’art

**Objectifs** L’une des premières problématiques abordées dans la robotique est la captation d’informations utiles à la mission. Dans le cadre de la proactivité, il est nécessaire d’avoir des informations concernant les utilisateurs présents dans l’environnement du robot. Pour cela, les capteurs utilisés furent ceux du robot Waldo, c’est-à-dire la caméra Realsense et la tablette Android. La tablette n’ayant aucune capacité pour capter des informations en amont de l’interaction, elle ne fut pas utilisée dans l’algorithme de suivi. Il s’agit donc de faire de la vision par ordinateur pour accomplir la prise d’informations en continu. De nombreuses informations peuvent être extraites dans le cadre de la proactivité, mais les plus intéressantes concernent celles obtenues en observant les humains. Il s’agit donc dans un premier temps de pouvoir les repérer dans le flux vidéo. De nombreuses recherches ont été faites sur le suivi de personnes en vision par ordinateur.

Les méthodes les plus fréquentes se basent sur une différenciation par rapport au fond pour permettre la détection de blobs [6, 1], exploitent des algorithmes d’apprentissage profond comme OpenPose<sup>1</sup>, utilisent des marqueurs [14], des systèmes avec plusieurs caméras, ou encore les algorithmes de détection de la Microsoft Kinect<sup>2</sup> [12].

Dans le cadre du projet certains critères concernant l’algorithme de suivi furent jugés nécessaires (dans l’ordre d’importance) :

- Il doit pouvoir fonctionner en temps réel ;
- Il doit pouvoir suivre plusieurs personnes à la fois ;
- Il ne doit pas nécessiter la présence d’un équipement de détection tiers ;
- Il doit pouvoir récupérer la position spatiale courante des personnes suivies ;
- Il doit estimer l’orientation globale, voire locale, de ces personnes ;

---

1. <https://github.com/CMU-Perceptual-Computing-Lab/openpose>

2. <https://developer.microsoft.com/fr-fr/windows/kinect>



Figure 1.1 – fonctionnement d'OpenPose

**Solutions et évolutions** L'algorithme OpenPose<sup>3</sup> précédemment cité fut celui sélectionné car il remplit directement les trois premières conditions [7, 8, 26, 31]. De plus, sa détection dans l'image de points d'intérêts (voir partie 1.2) permet de remplir les deux dernières conditions à l'aide de traitements supplémentaires (voir partie 1.3).

Toutefois, ces traitements supplémentaires font face à une problématique récurrente des algorithmes de suivi : les occlusions. Bien que certains algorithmes résistent aux occlusions partielles et aux occlusions totales sous certaines conditions - par exemple en inférant un modèle cinématique de chaque personne suivie [1, 6] -, les occlusions totales non négligeables dans le cas d'interactions Homme/Machine peuvent toujours entraîner des erreurs. Une piste souvent exploitée est de coupler le suivi avec une détection de caractéristiques biométriques propres à chacun (e.g.. empreinte digitale, iris, structure faciale, ...). Une méthode consiste à extraire des descripteurs pour chaque personnes avant occlusion [1], mais les descripteurs ne sont pas invariants aux changements chromatiques et de luminosité, et n'apportent donc pas la robustesse voulue.

La solution fut de coupler l'algorithme de détection et suivi précédent avec un algorithme d'identification de visage (voir partie 2).

## 1.2 Fonctionnement

La caméra envoie par le biais de ROS les flux d'images couleur et de profondeur. L'algorithme de détection de « squelettes » OpenPose est appliqué sur les images couleur.

Les « squelettes » sont ensuite localisés dans l'espace à l'aide de l'image de profondeur et des matrices de projection de chaque caméra (couleur et profondeur). L'algorithme de suivi (voir annexe B) vient ensuite mettre à jour ses connaissances sur l'environnement avec les résultats obtenus image par image. Cet algorithme :

<sup>3</sup>. implémentation sous ROS maintenue à jour régulièrement sur [https://github.com/firephinx/openpose\\_ros](https://github.com/firephinx/openpose_ros)

- stocke les informations spatiales enregistrées des personnes suivies dans une mémoire sur le long terme, appelée « Mémoire de Personnes » (MP) ;
- envoie vers l’algorithme d’identification les personnes suivies.

La MP se met à jour à chaque nouvelle information. Elle contient donc l’ensemble des informations que peut nécessiter une autre application tel que l’algorithme de comportement proactif.

## 1.3 Critère de suivi

L’algorithme OpenPose traite les images sans tenir compte des images précédentes. Il fut donc nécessaire d’instaurer un suivi gérant le déplacement spatial et l’écart temporel entre les images successives. De plus les résultats obtenus par OpenPose sont parfois bruités et sensibles à plusieurs facteurs : En présence de multiples personnes, l’algorithme peine à créer le bon nombre de composantes, et peut les associer aux mauvais « squelettes ». Similairement OpenPose a tendance à surestimer le nombre de composantes dans une image de trop bonne qualité. Enfin, certains éléments de l’environnement créent des faux positifs, et ce malgré l’absence d’une quelconque forme humanoïde.

Il fut donc nécessaire de fiabiliser les résultats grâce à la gestion temporelle précédemment énoncée.

### 1.3.1 Suivi spatial des « squelettes »

**Positionnement 3D** L’estimation des positions 3D de chaque point d’intérêt ne pouvant pas être faite dans l’algorithme OpenPose, la transposition des coordonnées en pixels aux coordonnées spatiales se fait à l’aide de la connaissance de la disposition spatiale des caméras par rapport à son environnement. En effet, les matrices de projection sont connues, il est donc possible d’estimer la correspondance entre une image couleur et son image de profondeur qui lui est temporellement corrélée. Ces calculs se font grâce à certains outils disponibles sur ROS<sup>4 5 6</sup>. Toutefois, l’estimation de la profondeur est bruitée, bien qu’un filtre médian sur la profondeur enlève une forte partie du bruitage. En effet les points d’intérêts extrêmes (oreilles, mains, ...) présentent tout de même une forte variance. En réponse, un second filtre médian appliqué à l’ensemble des profondeurs associées aux composantes du « squelette » permît de projeter ces dernières à la profondeur médiane et de supprimer la variance présentée précédemment. Il s’agit de ce fait d’une limite du système car l’estimation de l’orientation de la personne détectée en devint plus compliquée, voire impossible. Ce problème est rappelé dans les possibilités d’améliorations de la partie 5.3.

**Suivi spatial** Une des caractéristiques d’OpenPose est son fonctionnement en deux étapes, la détection de points d’intérêts dans l’image et leurs associations pour recréer les « squelettes »

---

4. [http://wiki.ros.org/image\\_geometry](http://wiki.ros.org/image_geometry)

5. [http://wiki.ros.org/tf2\\_ros](http://wiki.ros.org/tf2_ros)

6. [http://wiki.ros.org/message\\_filters#Time\\_Synchronizer](http://wiki.ros.org/message_filters#Time_Synchronizer)

précédemment présentés. Cela mène donc à la possibilité de « squelettes » incomplets, et à des associations erronées de membres en cas de proximité entre deux personnes. La solution apportée fut d'appliquer entre chaque image successive un algorithme d'association glouton basé sur une métrique de sommes pondérées de distances entre points d'intérêts communs.

En effet, à l'arrivée d'une nouvelle image, une collection de positions 3D de points d'intérêts, chacun associé à un « squelette », est calculée. Il s'agit donc de comparer ces « squelettes » avec les « squelettes » trouvés dans l'image précédente, et estimer si, dans le champ de vue de la caméra, une personne est apparue, a disparu ou s'est déplacée. Pour estimer cela, le choix fut de calculer la distance spatiale selon la norme euclidienne entre les « squelettes » de l'image précédente et ceux de la nouvelle image. Toutefois, les « squelettes » ne sont pas forcément composés des mêmes points d'intérêts. Il fut donc nécessaire de définir les positions à prendre en compte. Les possibilités furent :

- Le calcul des barycentres 1.1, puis les distances euclidiennes entre barycentres 1.2 ;
- les distances euclidiennes entre composantes communes 1.3, puis la moyenne de ces distances 1.4.

$$b_A = \frac{\sum_{k \in A} (w_k \times p_k)}{\sum_{k \in A} w_k} \quad (1.1)$$

$$dist_{A,B} = \|b_B - b_A\|_2 \quad (1.2)$$

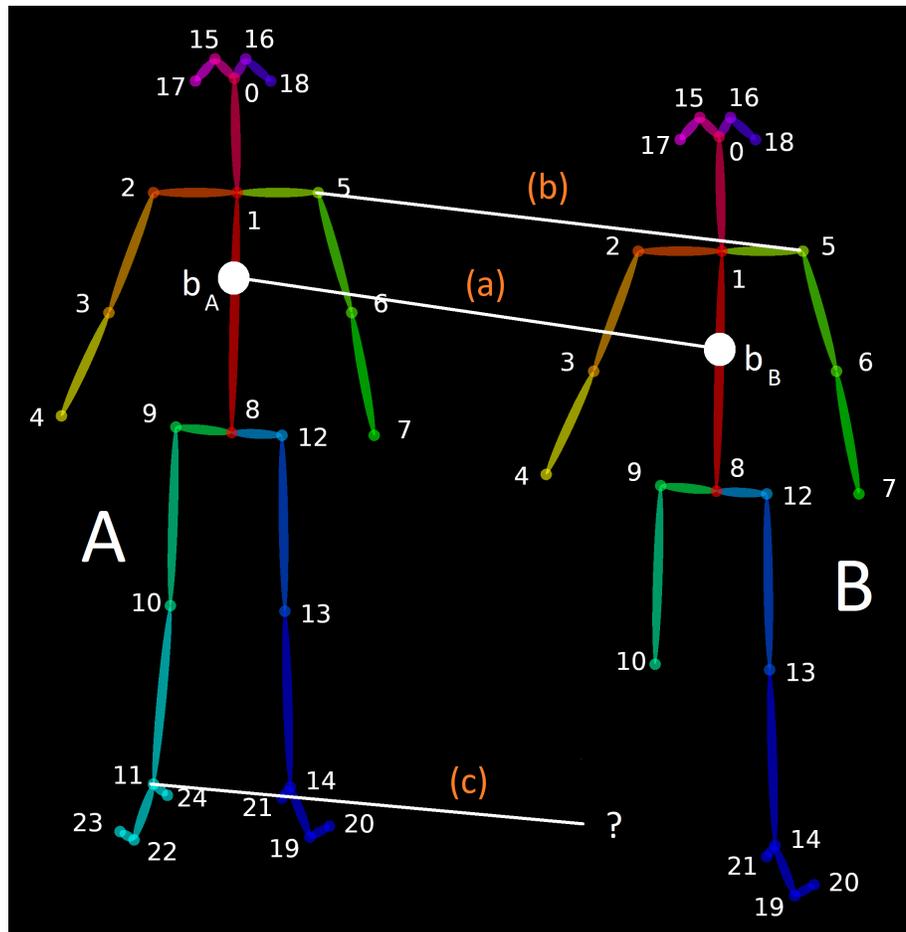
$$dist_{k \in (A \cap B)} = \|p_{k_B} - p_{k_A}\|_2 \quad (1.3)$$

$$dist_{A,B} = \frac{\sum_{k \in (A \cap B)} (w_k \times dist_{k \in (A \cap B)})}{\sum_{k \in (A \cap B)} w_k} \quad (1.4)$$

En notant  $w_k$  le poids associé à la composante  $k$ ,  $p_k = \begin{Bmatrix} x_k \\ y_k \\ z_k \end{Bmatrix}$  la position de la composante  $k$

dans un repère cartésien,  $b_A = \begin{Bmatrix} x_A \\ y_A \\ z_A \end{Bmatrix}$  la position du barycentre de l'objet  $A$ ,  $dist_{k \in (A \cap B)}$  la distance entre la composante  $k$  de l'objet  $A$  et celle de  $B$ , et  $dist_{A,B}$  la distance entre les objets  $A$  et  $B$ .

Les expérimentations sur la première méthode montrèrent une grande variance du posi-



**Figure 1.2** – Calcul de la distance entre les « squelettes » A et B. les composantes sont numérotées de 0 à 24.  $b_A$  et  $b_B$  sont les barycentres 1.1 des deux « squelettes ». (a) est la distance entre les barycentres 1.2, (b) est la distance  $dist_{5 \in (A \cap B)}$  et (c) est un exemple de composante ne se trouvant que sur l'un des « squelettes ».

tionnement des barycentres, car l'apparition/la disparition d'une composante pouvait tirer le barycentre dans sa direction. Une pondération plus importante des points centraux (cou, nez) diminue l'impact des apparitions et disparitions d'autres points, mais rend l'application très instable en leurs absences.

La deuxième méthode n'a pas cette variance, et n'a donc pas d'impact négatif sur le calcul de la distance globale. Elle nécessite toutefois une bonne estimation préalable des positions des points d'intérêts, mais une pondération plus forte sur les points centraux assure, en cas de leur présence, une meilleure estimation des distances. La seule limite est la situation où il n'y a aucune composante commune empêchant ainsi toute comparaison de distance, problème jugé négligeable en partant du postulat qu'une même personne a forcément une composante commune d'une image à la suivante tant que cette personne reste dans le champ de vue.

Une fois la matrice de distance (anciens « squelettes » X nouveaux « squelettes ») calculée, un algorithme glouton s'occupe de l'association. Celui-ci associe les « squelettes » les plus

proches par ordre de distances croissantes (selon la métrique définie précédemment). Il n'y a pas d'association si les distances restantes dépassent un certain seuil de distance paramétrable, ou s'il n'y a plus personne à associer.

Chaque « squelette » possède un identifiant qui lui est attribué quand il est détecté pour la première fois, cet identifiant est transmis à son potentiel prochain « squelette » associé.

### 1.3.2 Suivi temporel des « squelettes »

Le suivi spatial permet dans le cas d'une détection idéale d'associer correctement les « squelettes » par images successives. Néanmoins, OpenPose est un algorithme par apprentissage et est donc sujet aux faux positifs et aux faux négatifs. Cela entraîne des détections ponctuelles erronées (appelées par la suite « fantômes ») et des absences de détection de « squelettes » sur quelques images disparates (appelées par la suite « clignotements »).

La solution choisie fut de coupler le suivi spatial avec un suivi temporel. Un « squelette » passe d'abord dans une période d'approbation. Si ce « squelette » peut être suivi spatialement sans « clignoter » pendant une certaine période de temps paramétrable, il passe à l'état « approuvé ». Dans le cas contraire, il est considéré comme un « fantôme ».

Une fois le statut « approuvé » atteint, et tant que la période de temps de disparition dans son « clignotement » ne dépasse pas une période de temps seuil paramétrable, le « squelette » est toujours considéré comme présent. Dans le cas d'une disparition trop longue, le « squelette » est considéré comme disparu et n'est plus suivi. La position du « squelette » étant inconnu pendant son « clignotement », sa dernière position détectée est gardée en mémoire.

# Chapitre 2

## Algorithme d'identification

### 2.1 Objectifs

Comme expliqué précédemment, il est nécessaire dans un système de suivi robuste d'avoir une caractéristique unique permettant de différencier les instances suivis. Dans le cas présent il s'agit de faire une validation du suivi de personnes par identification faciale.

Le choix de la technologie de reconnaissance faciale s'est porté vers l'application Cobalt développée par une équipe rennaise d'Orange et mise à disposition aux autres projets via une API dans le *Cloud*. Les fonctionnalités de cette application sont expliquées dans la partie 2.4.

Les calculs étant déportés sur le réseau, le temps de traitement rend inapplicable la reconnaissance faciale sur l'entièreté du flux d'images. De plus, la reconnaissance faciale utilisée est sensible à la qualité de l'image étudiée. La solution fut de réduire le nombre d'appels à Cobalt grâce à des critères temporels, qualitatifs et spatiaux. Cela permet également de réduire la marge d'erreur d'identification.

### 2.2 Fonctionnement

A chaque itération l'algorithme de suivi juge l'intérêt de l'image reçue d'une part par la présence ou non de visages et d'autre part par la qualité de perception de ces derniers (voir partie 2.3). Un score d'intérêt d'identification du « squelette » associé à chaque visage varie selon ce jugement, en suivant la formule 2.1. Si le score atteint 100%, il est comparé à la base de données de visages (voir partie 2.4.1), et la MP est mise à jour. Si la personne est inconnue, une tentative de création de modèle est effectuée (voir partie 2.4.2).

$$s_k = \begin{cases} s_{k-1} - \frac{\Delta T}{c_{down}} & \text{si le visage est caché ou jugé de mauvaise qualité} \\ s_{k-1} + \frac{\Delta T}{c_{up}} & \text{sinon} \end{cases} \quad (2.1)$$

Avec  $s_k \in [0, 1]$  le score en pourcentage à l'itération  $k$ ,  $(c_{down}, c_{up})$  les coefficients d'évolution du score, et  $\Delta T$  le temps entre chaque itération d'identification.

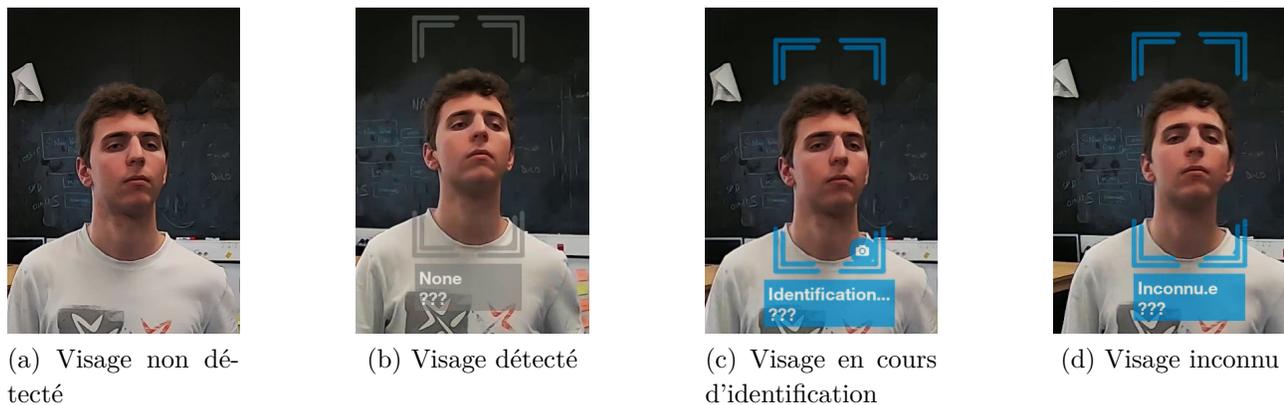


Figure 2.1 – Évolution du statut

## 2.3 Critère de qualité

Les points d'intérêts propres aux visages repérés par OpenPose permettent d'estimer la présence d'un visage sur l'image. Toutefois OpenPose est très robuste face à la mauvaise qualité d'image et a parfois tendance à extrapoler les points d'intérêts. L'objectif étant d'obtenir des modèles de visages fiables, il est nécessaire d'avoir une sécurité contre les extrapolations et les images de mauvaise qualité. Le problème d'extrapolation fut simplement corrigé par une règle nécessitant l'ensemble des points du visage (oreilles, yeux, nez, cou), ce qui garantit que ce dernier fait bien face à la caméra. Une supervision humaine a posteriori permît de valider ce critère de positionnement. La seconde problématique nécessita d'effectuer une estimation de la qualité d'une image, ce qui demanda une réflexion plus complète.

## 2.4 Critère de reconnaissance

### 2.4.1 Reconnaissance faciale

La reconnaissance faciale - telle qu'utilisée par l'application Cobalt - fait appel à des réseaux de neurones convolutifs estimant 70 points de repère à partir d'un visage dans une image. La comparaison de ces points avec ceux établis dans une base de données de modèles composés de ces mêmes points permet de s'assurer de l'identité de la personne. Cela permet donc de gérer les problèmes d'occlusion totale et de sortie/entrée d'une personne dans le champ de vision de la caméra évoqués dans la partie 1.1.

### 2.4.2 Création de modèle

Dans le cas où l'identification n'a pas trouvé de correspondance avec un modèle déjà présent dans la base de données - ce qui est le cas à la première rencontre avec une personne - il s'agit de vérifier s'il est possible de créer un modèle. Une fonctionnalité de Cobalt permet de créer

un modèle de visage à partir d'une sélection de visages (les points de repère présentés dans la partie 2.4.1). Le modèle utilisé gagne en robustesse si la sélection est suffisamment grande (un minimum de cinq visages). L'application nécessite donc de rassembler un corpus d'au moins cinq visages d'une même personne.

Cela est accompli par une autre fonctionnalité de Cobalt. Celle-ci permet de créer des groupes de visages similaires - donc de la même personne - parmi une liste de visages qui lui a été transmise. Ainsi, à la réception d'un nouveau visage inconnu lors de l'identification, l'application développée crée les corpus de visages à partir de tous les visages enregistrés encore inconnus, et crée un modèle pour chaque corpus suffisamment gros.

## 2.5 Discussion sur l'approche globale

Un couplage entre l'algorithme de suivi (voir partie 1) et l'algorithme d'identification (voir partie 2) fut la méthode utilisée pour obtenir les informations nécessaires à l'implémentation du comportement proactif voulu. L'efficacité de la méthode est abordée dans la partie 4. De nouvelles pistes d'améliorations sont présentées dans la partie 5.3.



# Chapitre 3

## Algorithme de proactivité

Les algorithmes de suivi et de reconnaissance précédemment présentés furent proposés comme base de travail pour alimenter des algorithmes d'interactions sociales. Il convient toutefois d'effectuer dans un premier temps une étude des motivations à initier de telles interactions. Il s'agit donc de définir le principe de proactivité, d'en étudier les travaux associés et de proposer une implémentation en accord avec le contexte d'utilisation.

### 3.1 Proactivité

#### 3.1.1 Formalisation

La proactivité (ou pro-activité) est un néologisme venant de l'anglais qui se définit par la prise d'action en prévision d'une future situation, ou pour amener à cette situation<sup>1</sup>. L'implémentation de ce genre de comportement sur un robot permet :

- De se mettre dans une situation poussant à la demande d'accomplissement d'une mission, sans déclencheur particulier. Par exemple, En l'absence de personne dans l'environnement du robot, celui-ci décide de partir à la recherche d'utilisateur. Cette situation n'est pas étudiée ici ;
- D'accomplir une mission si jugé nécessaire, avec déclencheur mais sans demande explicite. Par exemple, le robot voit une personne le regarder de manière soutenue, et décide en conséquence d'engager la discussion.

Dans les deux cas, cela peut soit préciser un besoin, soit en créer un. Les intérêts sont :

- un gain de temps dans les interactions, et donc dans l'efficacité du service ;
- une apparente serviabilité et fiabilité accrue, et donc un capital sympathie plus grand [30].

---

1. <https://sites.google.com/site/profsharonparker/proactivity-research/what-is-proactivity>

### 3.1.2 Limites

De la formalisation découle le fait qu'un comportement proactif peut être totalement indépendant de tout accord explicite préalable de l'utilisateur, ce qui peut être anxiogène. Les travaux sur les voitures autonomes tel que [25] ont étudié le sujet, sans pour autant avoir trouver une solution à cette problématique. L'acceptation de comportements automatiques, et plus particulièrement de la robotique, ne sera pas traitée ici.

Concernant la proactivité dans les robots sociaux, leur côté intrusif et leurs comportements artificiels peuvent devenir plus acceptables, en s'inspirant de travaux comme [24].

## 3.2 État de l'art

- De nombreuses études sont consacrées à la psychologie comportementale, et certaines peuvent servir de référence pour la robotique sociale. Par exemple, Y. Xing et al. présente une étude du comportement des conducteurs, et a établi dans un premier temps qu'il y a une différenciation des comportements en fonction de si ceux-ci sont en réaction à des stimulus, ou guidés par des objectifs [32]. Il en déduit aussi que les déplacements de tête et du regard reflètent le processus de pensée associé à la récupération d'informations.
- Une architecture classique dans le domaine des agents est le Belief-Desire-Intention (BDI)<sup>2</sup>. Il s'agit d'une représentation du système de raisonnement s'appuyant sur l'application de comportements prédéfinis en fonction de la perception de l'environnement. Il existe des variantes appliquées à la robotique (e.g. PROFETA [16]<sup>3</sup>), et plus particulièrement à la robotique sociale (e.g. robots compagnons [19]).
- Plusieurs recherches sur la détection d'intentions dans le cas de la production de tâches en coopération. Une majeure partie se base sur des machines à états ou des réseaux bayésiens dont les transitions sont déclenchées par l'utilisateur ou par le changement spatial de l'environnement (entre autres [2, 3, 20, 21]), mais il existe d'autres approches plus particulières. J. Han et al. apprennent des hiérarchies entre les comportements, les objets, la disponibilité (de l'anglais *affordance*) de ces derniers, et cela en adoptant des points de vue différents sur l'environnement de fonctionnement [17]. S. Trick et al. étudie la voix, les gestes, le regard et la position des objets d'interaction, puis effectue une fusion de classifieurs pour réduire la marge d'erreur dans l'estimation de l'intention [29]. Le variété de tâches à accomplir n'étant pas toujours définissable à l'avance, des travaux (comme [4, 22]) ont été effectués pour apprendre dynamiquement de nouveaux comportements dans le cas où un comportement n'a jamais été rencontré. Enfin, certains travaux prennent en compte les possibles changements d'intentions lors de la réalisation

---

2. Plus d'informations sur [https://tams.informatik.uni-hamburg.de/lehre/2017ws/seminar/ir/doc/slides/MassimoInnocenti-Belief\\_Desire\\_Intention.pdf](https://tams.informatik.uni-hamburg.de/lehre/2017ws/seminar/ir/doc/slides/MassimoInnocenti-Belief_Desire_Intention.pdf), <https://msdn.microsoft.com/en-us/magazine/mt848706.aspx> et [http://turing.cs.pub.ro/auf2/html/chapters/chapter2/chapter\\_2\\_2\\_2.html](http://turing.cs.pub.ro/auf2/html/chapters/chapter2/chapter_2_2_2.html)

3. Une implémentation de PROFETA sur <https://github.com/corradosantoro/profeta>

d'une tâche [21].

- De nombreux papiers de recherche sont focalisés sur la détection d'émotions et certains d'entre eux proposent d'effectuer une réaction en adéquation à ces émotions[19, 23]. Cela mène à une réflexion sur le lien entre l'état d'esprit et le comportement, et donc sur son impact sur la proactivité.
- K. Drejing et al. définit la motivation à interagir par l'intensité de l'engagement et les comportements spécifiques associés [13]. Il faut donc tempérer tout comportement proactif selon le degré d'intérêt de la personne cible.
- Finalement, certains travaux prônent une proactivité plus agressive : Y. Tamura et al. détecte de manière objective les zones d'intérêts dans une image en se fondant sur l'observation du regard et des mains d'un magicien ainsi que de la saillance (de l'anglais salience) de l'image dans le but d'implémenter dans un robot un comportement manipulateur [27]. Similairement, M. M. Hoque et al. détecte l'orientation du visage et du regard pour définir un comportement qui attire le plus possible l'attention [18].

### 3.3 Implémentation

Comme présenté précédemment, la proactivité dans l'interaction Homme/Machine est un sujet assez complexe et relativement nouveau dans le domaine de la recherche en robotique. L'objectif est ici d'initier une interaction sociale dans la continuation des travaux sur le suivi de personnes (voir partie 1). Dans le cadre d'une gestion de projet Agile (voir partie ) une première implémentation naïve fut expérimentée.

#### 3.3.1 Fonctionnement

**Score « d'intérêt »** Pour chaque personne visible (voir partie 1.2) et identifiée (possédant un modèle de visage comme décrit dans la partie 2.4.2) un score « d'intérêt » lui est associée. Dans le cas où il s'agit de la première identification de cette personne le score est initialisé à 0. Mais si la personne a déjà été croisée auparavant, le score s'initialise selon la formule 3.1. Cela permet de gérer les entrées et sorties rapides du champ de vision de la caméra, et les occlusions complètes courtes.

Une fois initialisé, le score évolue selon la formule 3.2.

$$i_0 = \begin{cases} i' - \frac{\Delta T'}{b_0} & \text{si la personne identifiée avait déjà été vue auparavant} \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

Avec  $i_k \in [0, 1]$  le score en pourcentage à l'itération k,  $i'$  le score « d'intérêt » du précédent suivi de cette personne,  $b_0$  le coefficient temporel d'évolution du score quand la personne n'est

pas visible, et  $\Delta T'$  le temps passé depuis la dernière détection.

$$i_k = \begin{cases} i_{k-1} + \frac{\Delta T}{b_{up}} & \text{si l'on voit cou, nez, oreilles} \approx \text{"la personne regarde le robot"} \\ i_{k-1} - \frac{\Delta T}{b_{down}} & \text{sinon} \end{cases} \quad (3.2)$$

Avec  $(b_{down}, b_{up})$  les coefficients d'évolution du score, et  $\Delta T$  le temps entre chaque itération de détection.

**Suivi de la tête** Mise à part l'évolution du score « d'intérêt », un suivi de personnes est effectué par la tête du robot. La tête cible la personne (son visage si possible) possédant le plus haut score, et la suit jusqu'à la perdre de vue. La seule exception concerne une implémentation d'une première interaction. Le robot Waldo entame un dialogue avec la personne la plus intéressante selon le critère précédemment énoncé, ce qui déclenche le suivi de cette personne par la tête.

### 3.3.2 Limites

Le fonctionnement présenté est assez limité. En effet l'utilisation d'un score « d'intérêt » n'est appuyée par aucune utilisation ultérieure scientifiquement éprouvée, et ses paramètres d'évolution furent établis expérimentalement. Une première étape serait de fixer ces derniers plus rigoureusement, en utilisant par exemple un algorithme d'apprentissage. Il serait bien sûr plus intéressant d'implémenter une des architectures présentées dans l'état de l'art de la partie 3.2.

# Chapitre 4

## Expérimentations

### 4.1 Cas d'utilisation

L'objectif du présent document est de démontrer le possible intérêt de la proactivité dans le cadre d'un service qu'Orange pourrait possiblement proposer à l'avenir, tout en effectuant une veille technologique dans le domaine de la robotique. Un scénario récurrent dans ceux proposés au sein de l'équipe fut de réaliser une collaboration entre un robot humanoïde accueillant des visiteurs, et une base robotique mobile guidant les visiteurs jusqu'à une destination voulue. Ce scénario a pour avantage de mettre en avant l'ensemble des compétences nécessaires à la robotique, et propose de surcroît un service en adéquation avec le domaine d'activité d'Orange.

L'expérimentation ne concerne que la partie accueil du service. Le robot Waldo fut placé à l'accueil des locaux de l'équipe robotique. Une explication du démonstrateur fut placée à côté de celui-ci, laissant la liberté à chacun de participer ou non à l'expérience. Le graphique en annexe C représente le cas d'utilisation. Il permet à la fois de juger les différents éléments présentés précédemment sur leurs bons fonctionnements conjoints, ainsi qu'obtenir le ressenti vis-à-vis du comportement du robot.



Figure 4.1 – Interface sur Waldo

## 4.2 Résultats

L'expérimentation permît de récupérer un ensemble d'environ 1000 sessions de personnes suivies. Chaque session contient différentes informations concernant le suivi, la reconnaissance faciale et les interactions. À l'aide d'une interface créée en interne, les données récupérées lors de l'expérimentation furent analysées manuellement. La visualisation des événements et sessions permet de comprendre rapidement des comportements du système.

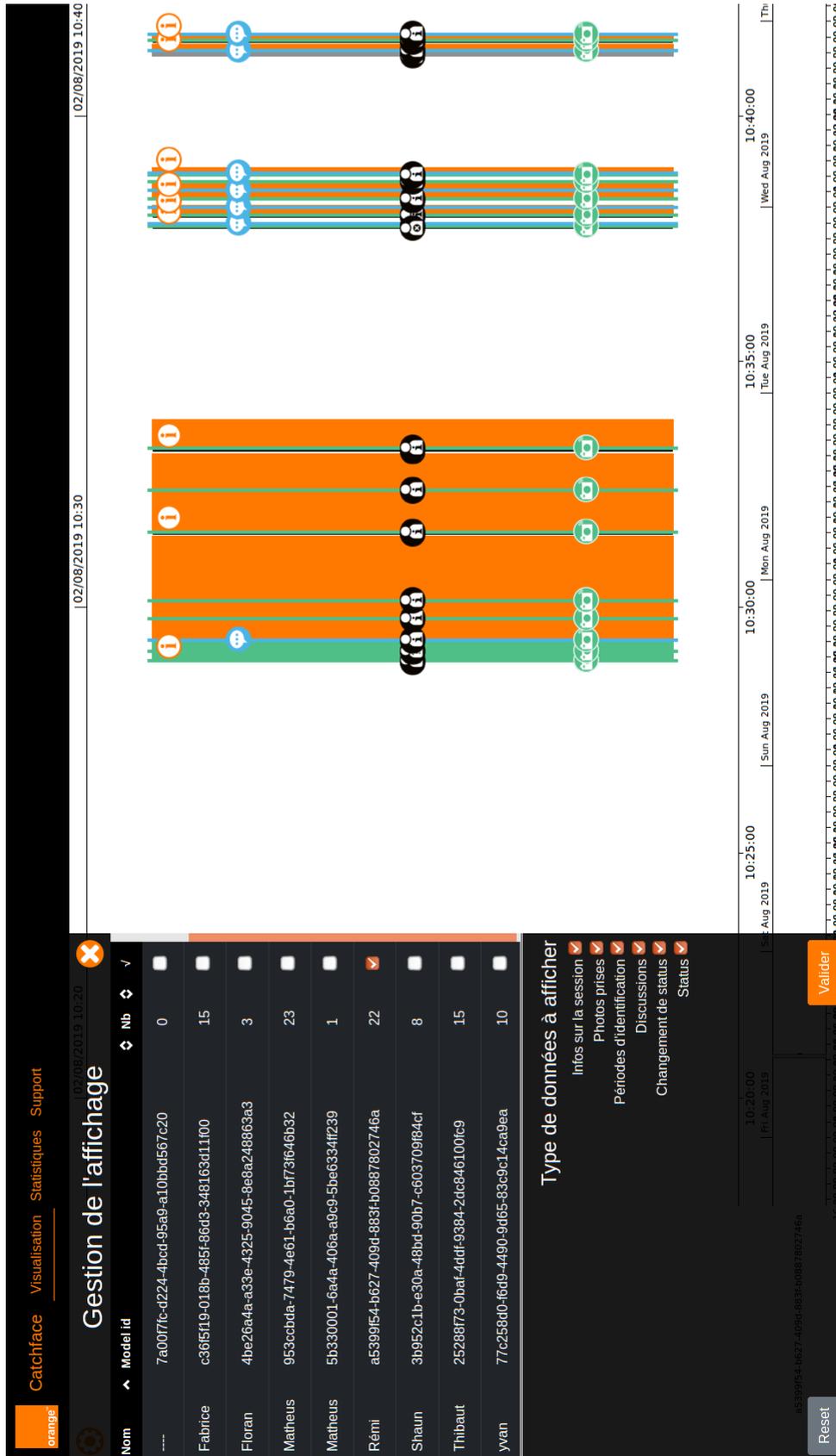


Figure 4.2 – Interface d'affichage des données

Les sessions de personnes suivies associées à chaque modèle de visage correspondent bien à la même personne. Il n'y a donc pas de faux positifs. En revanche, certaines sessions n'ont pas été associées bien que les images prises correspondaient à des modèles de visages connus. Les faux négatifs ainsi repérés sont en faible nombre.

Dans les personnes identifiées (modèle de visage créé), certains n'attendaient pas l'étape d'identification par l'interface, ou ne la finalisait pas.

### 4.3 Interprétation

L'absence de faux positifs tend à prouver le bien fondé de la méthode utilisée pour gérer les cas litigieux. Il faut toutefois prendre en compte qu'il fut rare de voir de multiples personnes en même temps dans l'environnement d'expérimentation, ce qui réduit considérablement les problématiques d'occlusions et d'appropriations d'identités.

Presque tous les faux négatifs peuvent être imputés à des images de mauvaise qualité ou à des visages partiellement occultés, ce qui fut sciemment jugé irrecevable.

Concernant les interactions abandonnées, selon le cas cela peut s'expliquer par plusieurs raisons :

- L'utilisateur fut inattentif ;
- L'interface d'interaction manqua d'instructions assez explicites ;
- La démarche proactive fut trop peu expressive ;
- L'utilisateur fut réticent à s'identifier.

Les points relatifs à la réticence et au manque d'expressivité méritent un approfondissement dans de futurs travaux, comme présenté dans la partie 5.3.

# Chapitre 5

## Conclusion et perspectives

### 5.1 Conclusion

Dans ce rapport une proposition d'architecture permettant un comportement proactif est présentée. Elle s'appuie sur un algorithme de détection de personnes dans un flux vidéo 2D nommé OpenPose réappliqué dans un flux vidéo 3D, d'un algorithme de suivi image par image utilisant la disposition spatiale fiabilisée par une composante temporelle, et d'un mécanisme centré sur l'identification de visages. Un état de l'art de la proactivité dans la robotique sociale est présenté et une première implémentation est expérimentée.

Bien que perfectible, l'implémentation effectuée offre des résultats à la hauteur des attentes de l'équipe, ainsi qu'un retour d'expérience riche grâce aux participants.

### 5.2 Mémorisation et données personnelles

Dans le contexte du Règlement Général sur la Protection des Données (RGPD, ou GDPR en anglais), il fut nécessaire de concevoir l'architecture des différents algorithmes en y ajoutant une réflexion sur la gestion des données. Ainsi, les données récupérées lors du suivi et de la reconnaissance sont associées à une clé unique. Cette clé est générée lors de la création de modèle (voir partie 2.4.2), et est donc associée à un visage. Toutefois, les images correspondantes à chaque modèle sont gérées par l'application Cobalt, tandis que les informations provenant du suivi sont gérées en local. Il y a donc une séparation (améliorable) des données pouvant assurer le respect des réglementations en vigueur. De plus l'espace de stockage des données du suivi est paramétrable, ce qui permettrait de privatiser ces données et d'en avoir le contrôle total. Ainsi, dans le cas d'un service à domicile, les utilisateurs n'auraient pas à se soucier du devenir de leurs informations personnelles.

### 5.3 Avenir du projet

Outre les possibles améliorations du comportement proactif présentées précédemment dans la partie 3.2, il existe d'autres pistes d'évolution du service à explorer :

- Le positionnement spatial des « squelettes » est imparfait du fait du bruit de mesure présenté dans la partie 1.3.1. Fiabiliser la détection de la profondeur permettrait une meilleure estimation de la posture humaine et donc une meilleure détection de l'intention.
- De nombreux travaux [27, 33] s'accordent à définir le regard comme l'un des points les plus importants pour identifier l'intention d'une personne. Implémenter une détection de direction de regard améliorerait grandement la précision du système.
- Le robot Waldo actuellement utilisé est capable de se déplacer précisément dans son environnement. Dans l'éventualité d'un service nécessitant un déplacement, la méthodologie d'approche est à développer comme proposée dans [9].
- Dans le cas où le visage n'est pas disponible (par manque ou par choix), d'autres caractéristiques biométriques peuvent être utilisées. Par exemple, W. Chi et al. utilise la stature pour identifier l'utilisateur [11]. Un vêtement intelligent pourrait aussi être utilisé [10], changeant en partie le paradigme adopté dans ce papier, mais offrant une plus grande variété d'informations.
- La proactivité développée ici ne prend en compte que l'engagement d'une interaction, mais il serait tout aussi intéressant de développer dans un second temps la méthode de désengagement, en s'inspirant par exemple des travaux présentés dans [24].
- Au vu de l'impact du suivi de la personne par la tête de Waldo, il serait intéressant d'implémenter plus de comportements non-verbaux ou co-verbaux [5]. En effet le manque d'expressivité des plateformes robotiques peut entraîner une incompréhension de la volonté du robot par l'utilisateur, ce qui peut être amélioré (voire corrigé) par l'utilisation d'actionneurs comme un bras ou un affichage numérique. Il faudra toutefois adapter les comportements en fonction des limites de ces derniers, la morphologie robotique étant parfois mal adaptée [5, 28].
- Une réserve concerne l'utilisation de robots humanoïdes. En effet la volonté de mettre en œuvre le biomimétisme peut susciter le phénomène de « vallée de l'étrange » et de ce fait entraîner des inquiétudes, voire un rejet comme évoqué dans [15].

# Bibliographie

- [1] Roberto Arroyo, J. Javier Yebes, Luis M. Bergasa, Iván G. Daza, and Javier Almazán. Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert Systems with Applications*, 42(21) :7991 – 8005, 2015.
- [2] M. Awais and D. Henrich. Human-robot collaboration by intention recognition using probabilistic state machines. In *19th International Workshop on Robotics in Alpe-Adria-Danube Region (RAAD 2010)*, pages 75–80, June 2010.
- [3] M. Awais and D. Henrich. Proactive premature intention estimation for intuitive human-robot collaboration. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4098–4103, Oct 2012.
- [4] M. Awais and D. Henrich. Human-robot interaction in an unknown human intention scenario. In *2013 11th International Conference on Frontiers of Information Technology*, pages 89–94, Dec 2013.
- [5] P. Bremner and U. Leonards. Efficiency of speech and iconic gesture integration for robotic and human communicators - a direct comparison. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1999–2006, May 2015.
- [6] Marcus A. Brubaker, Leonid Sigal, and David J. Fleet. *Video-Based People Tracking*, pages 57–87. Springer US, Boston, MA, 2010.
- [7] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose : Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008, 2018.
- [8] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1611.08050, 2016.
- [9] D. Carton, A. Turnwald, K. Kehlrenz, D. Wollherr, and M. Buss. Proactive human approach in dynamic environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3320–3321, Oct 2012.
- [10] M. Chen, J. Zhou, G. Tao, J. Yang, and L. Hu. Wearable affective robot. *IEEE Access*, 6 :64766–64776, 2018.
- [11] W. Chi, J. Wang, and M. Q. Meng. Person verification based on skeleton biometrics by rgb-d camera. In *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, pages 671–676, Dec 2014.

- [12] W. Chi, J. Wang, and M. Q. . Meng. A gait recognition method for human following in service robots. *IEEE Transactions on Systems, Man, and Cybernetics : Systems*, 48(9) :1429–1440, Sep. 2018.
- [13] K. Drejing, S. Thill, and P. Hemeren. Engagement : A traceable motivational concept in human-robot interaction. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 956–961, Sep. 2015.
- [14] Moataz Eltoukhy, Shihab Asfour, Craig Thompson, and Loren Latta. Evaluation of the performance of digital video analysis of human motion : Dartfish tracking system. *International Journal of Scientific and Engineering Research*, 3, 01 2012.
- [15] Francesco Ferrari, Maria Paola Paladino, and Jolanda Jetten. Blurring human-machine distinctions : Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, 8(2) :287–302, Apr 2016.
- [16] Loris Fichera, Daniele Marletta, Vincenzo Nicosia, and Corrado Santoro. Flexible robot strategy design using belief-desire-intention model. In David Obdržálek and Achim Gottscheber, editors, *Research and Education in Robotics - EUROBOT 2010*, pages 57–71, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [17] J. Han, S. Lee, and J. Kim. Behavior hierarchy-based affordance map for recognition of human intention and its application to human-robot interaction. *IEEE Transactions on Human-Machine Systems*, 46(5) :708–722, Oct 2016.
- [18] M. M. Hoque, Q. D. Hossian, D. Das, Y. Kobayashi, Y. Kuno, and K. Deb. An empirical robotic framework for interacting with multiple humans. In *2013 International Conference on Electrical Information and Communication Technology (EICT)*, pages 1–5, Feb 2014.
- [19] Wafa Johal, Damien Pellier, Carole Adam, Humbert Fiorino, and Sylvie Pesty. A cognitive and affective architecture for social human-robot interaction. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, HRI’15 Extended Abstracts, pages 71–72, New York, NY, USA, 2015. ACM.
- [20] W. Y. Kwon and I. H. Suh. A temporal bayesian network with application to design of a proactive robotic assistant. In *2012 IEEE International Conference on Robotics and Automation*, pages 3685–3690, May 2012.
- [21] T. Liu, J. Wang, and M. Q. . Meng. Human robot cooperation based on human intention inference. In *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, pages 350–355, Dec 2014.
- [22] T. Liu, J. Wang, and M. Q. . Meng. Evolving hidden markov model based human intention learning and inference. In *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 206–211, Dec 2015.
- [23] Z. Liu, F. Pan, M. Wu, W. Cao, L. Chen, J. Xu, R. Zhang, and M. Zhou. A multimodal emotional communication based humans-robots interaction system. In *2016 35th Chinese Control Conference (CCC)*, pages 6363–6368, July 2016.

- [24] Silvia Rossi, Giovanni Ercolano, Luca Raggioli, Emanuele Savino, and Martina Ruocco. The disappearing robot : An analysis of disengagement and distraction during non-interactive tasks. pages 522–527, 08 2018.
- [25] K. E. Schaefer, R. W. Brewer, J. Putney, E. Mottern, J. Barghout, and E. R. Straub. Relinquishing manual control : Collaboration requires the capability to understand robot intent. In *2016 International Conference on Collaboration Technologies and Systems (CTS)*, pages 359–366, Oct 2016.
- [26] Tomas Simon, Hanbyul Joo, Iain A. Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. *CoRR*, abs/1704.07809, 2017.
- [27] Y. Tamura, S. Yano, and H. Osumi. Visual attention model for manipulating human attention by a robot. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5307–5312, May 2014.
- [28] Yusuke Tamura, Takafumi Akashi, , and Hisashi Osumi. Where robot looks is not where person thinks robot looks. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 21(4) :660–666, 2017.
- [29] Susanne Trick, Dorothea Koert, Jan Peters, and Constantin A. Rothkopf. Multimodal uncertainty reduction for intention recognition in human-robot interaction. *CoRR*, abs/1907.02426, 2019.
- [30] Dominique Vaufreydaz, Wafa Johal, and Claudine Combe. Starting engagement detection towards a companion robot using multimodal features. *CoRR*, abs/1503.03732, 2015.
- [31] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. *CoRR*, abs/1602.00134, 2016.
- [32] Y. Xing, C. Lv, H. Wang, H. Wang, Y. Ai, D. Cao, E. Velenis, and F. Wang. Driver lane change intention inference for intelligent vehicles : Framework, survey, and challenges. *IEEE Transactions on Vehicular Technology*, 68(5) :4377–4390, May 2019.
- [33] Q. Zhao, D. Tu, S. Xu, H. Shao, and Q. Meng. Natural human-robot interaction for elderly and disabled healthcare application. In *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 39–44, Nov 2014.



# Annexe A

## Actionlib enhanced

Lors de la mise en place des différents algorithmes sous ROS, le manque d'une brique d'échange de messages asynchrone robuste fut remarqué. En effet, le fait de ne pas faire certains traitements (e.g. identification des visages par l'API d'Orange) en parallèle nuit au fonctionnement global de l'application. La librairie Actionlib<sup>1</sup> permet l'implémentation d'une telle mécanique, mais requiert un fort effort d'adaptation du code. La surcouche « simpleActionLib » proposée permet une interface plus haut niveau de la librairie et plus imple d'utilisation mais fait défaut sur plusieurs points :

- Elle ne gère pas plusieurs clients pour un même serveur ;
- Elle ne gère qu'au maximum deux requêtes par serveur, préempte la requête en cours avec la dernière requête reçue ;
- Elle ne fait pas de suivi par identifiant des requêtes, il n'y a donc aucun moyen de savoir quelle requête correspond au retour de l'application.

La solution fut d'exploiter l'implémentation bas-niveau de la librairie et d'y apporter une gestion multIClient et par exécution concurrente de requêtes. Ces dernières sont donc traitées en parallèle, mais résolues chronologiquement grâce à leurs identifiants. Ainsi, il n'y a pas besoin du système de préemption proposé par la librairie utilisée. De plus, par souci d'expérience utilisateur, l'utilisation de la librairie développée est très similaire à la surcouche « simpleActionLib ». Le code source peut se trouver sur <https://github.com/fannibal/actionlib-enhanced>.

---

1. <http://wiki.ros.org/actionlib>



# Annexe B

## item-tracking

L’algorithme de suivi devant être indépendant de la technologie utilisée pour détecter les personnes, il fut jugé nécessaire de développer une brique technologique pouvant effectuer un suivi de n’importe quel type d’objet, tant que celui-ci se définit par un ou plusieurs points de repère identifiés. Par exemple, les « squelettes » détectés par OpenPose sont définis par des composantes correspondant à des parties de corps (e.g. oreille gauche, cou, ...) repérées dans un repère cartésien.

La librairie développée permet d’effectuer le suivi de plusieurs objets par détections successives, spatialisées et temporalisées de leurs composantes nommées. Cet algorithme de suivi est donc capable de suivre différents types d’objets.

Les étapes à chaque appel à la fonction sont :

- Estimation des barycentres des objets ajoutés ;
- Calcul des distances entre chaque objet ajouté et les objets précédemment suivis à partir de leurs composantes ;
- Estimation de l’évolution du suivi (ajout, mis à jour, retrait) par rapport aux distances et aux paramètres globaux.

Le fonctionnement de cet algorithme appliqué au suivi d’humain est décrit plus précisément dans la partie 1.3. Le code source peut se trouver sur [https://github.com/fannibal/item\\_tracking](https://github.com/fannibal/item_tracking).



## **Annexe C**

### **Logigramme du cas d'utilisation**

