

# Etat de l'art des technologies de reconnaissance de gestes

---

## Introduction :

L'essor des technologies d'intelligence artificielle a permis de proposer des applications adressant des besoins très spécifiques inatteignables par les paradigmes de la programmation classique. Ainsi, nous avons vu ces technologies s'intégrer de plus en plus à nos vies de tous les jours. Maisons connectées, outils d'accessibilité sur les téléphones mobiles, bracelets et montres connectés, sont quelques applications parmi d'autres qui proposent des fonctionnements adaptés à l'utilisateur en fonction de ses habitudes. Ces habitudes prennent une dimension plus concrète sous forme de flux de séries temporelles. Ces données sont étudiées par les algorithmes d'apprentissage pour proposer l'expérience la plus adaptée et la plus personnalisée en fonction de chaque utilisateur.

Dans ce document, on se propose d'étudier une fraction de ces applications, à travers l'exemple des applications destinées à la reconnaissance des gestes effectués par les utilisateurs. Dans un premier temps, nous allons situer le contexte technologique, avant de décrire les méthodes et les types de données utilisés pour reconnaître et recenser ces gestes. Ensuite, nous présenterons quelques cas d'utilisation de ces technologies, en dernière partie, on discutera quelques problématiques ainsi que les méthodes de résolution adoptées dans la littérature étudiée.

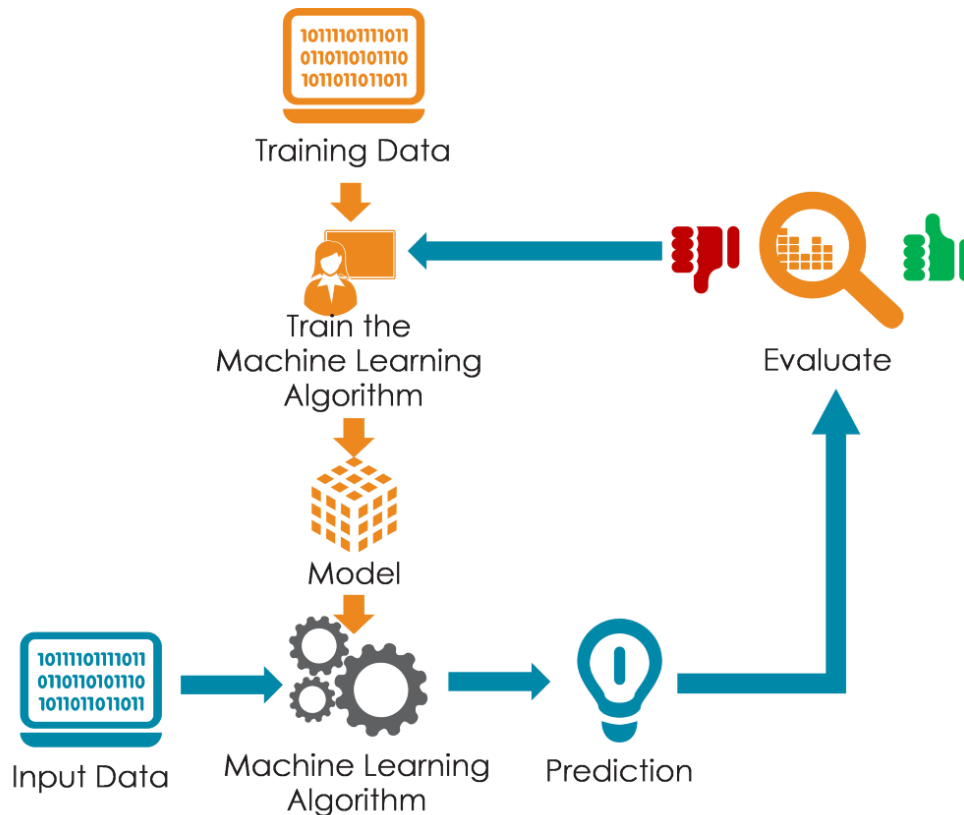
## Contexte technologique : AI / TinyML / Edge computing :

### 1. L'intelligence artificielle : une nouvelle approche de visualiser les données.

La démocratisation des algorithmes d'intelligence artificielle constitue l'un des événements qui ont le plus marqué l'évolution technologique. Ce phénomène est principalement dû à l'avancée technologique au niveau hardware et à la baisse des prix des composants de calcul tels que les processeurs graphiques. Cette baisse peut être tracée aux usages des GPUs dans le domaine du divertissement, et du gaming en particulier. Ainsi, ces composants offrant des capacités de calcul parallèles étaient accessibles, ce qui a permis une évolution très rapide du domaine de l'IA qui a vu le jour dans les années 1950. Les limitations rencontrées à l'époque en termes de tailles de données et faiblesse des composants de calculs présentaient ce qu'on appelle des "bottleneck" qui

ralentissaient considérablement la recherche dans ce domaine.

Le principe de fonctionnement des algorithmes d'intelligence artificielle n'étant plus secret à personne, il n'est pas aussi simple qu'il le semble. Il s'agit de concevoir des algorithmes permettant d'ajuster leurs sorties en fonction des entrées reçues. La conception de ces algorithmes est principalement basée sur des modèles statistiques combinés à une multitude de méthodes d'optimisation. Ces algorithmes ou "modèles" sont conçus suivant un enchaînement précis présenté sur le schéma<sup>1</sup> suivant :



Le schéma précédent permet de diviser le processus de conception en deux étapes: une étape d'entraînement et une étape de test. l'étape d'entraînement permet au modèle d'ajuster l'ensemble de ses paramètres pour adapter sa prédiction à la valeur réelle du résultat souhaité. Le processus d'ajustement est effectué différemment selon les modèles, les types d'apprentissage ( supervisé, non supervisé, clustering, ...) ainsi que la tâches auquel le modèle est destinée (classification, régression,...). La deuxième étape étant l'évaluation du modèle après son entraînement par un jeu de données qui n'a pas été utilisé pendant la phase précédente, afin de conclure sur la précision du modèle.

Ces algorithmes nécessitent des quantités considérables de données. Ce qui requiert une puissance de calcul importante pour l'entraînement, d'où l'apparition et l'évolution des technologies de calcul dans le cloud, car ces

<sup>1</sup> [https://cdn-images-1.medium.com/max/1000/1\\*E7sk9C5buY9KnBUDfl9fQ.png](https://cdn-images-1.medium.com/max/1000/1*E7sk9C5buY9KnBUDfl9fQ.png)

plateformes, hébergés dans des data centers, offrent des services permettant aux entreprises ainsi qu'aux particuliers d'avoir à leurs disposition, des machine

“virtuelles” leurs permettant d'avoir des puissance de calcul énorme, tout en se passant par les coûts d'achats et de maintenance de ces machines.

Les services mentionnés précédemment sont proposés avec un mode de facturation selon l'usage (en heures ou jours de calcul), avec la possibilité de réduire et d'augmenter la puissance de calcul en fonction du besoin.

L'usage des services cloud représente un grand saut technologique. Néanmoins, pour des applications nécessitant le transfert de données massifs de leurs sources aux datacenters avant de recevoir les résultats de calculs plusieurs contraintes font surface. Les problématiques majeures sont liées principalement à la sécurité des données, puisque les données transitent sur le réseau pour arriver à un datacenter, où ils seront stockés puis traités. Ce qui pose un risque majeur de sécurité dans le cas des intrusions ou hacks sur des datacenters, ce qui met en danger les données sensibles des utilisateurs et qui peut les mettre en danger dans d'autres. L'exemple le plus illustratif étant celui du fameux hack de l'une des voitures de Tesla<sup>2</sup>.

D'autres problématiques liées à la disponibilité des services sont abordées lors des problèmes<sup>3</sup> sur les datacenters hébergeant leurs services, d'autres liés à aux problèmes de latence dans des cas d'applications temps-réels.

## 2. Edge computing : une solution magique ?

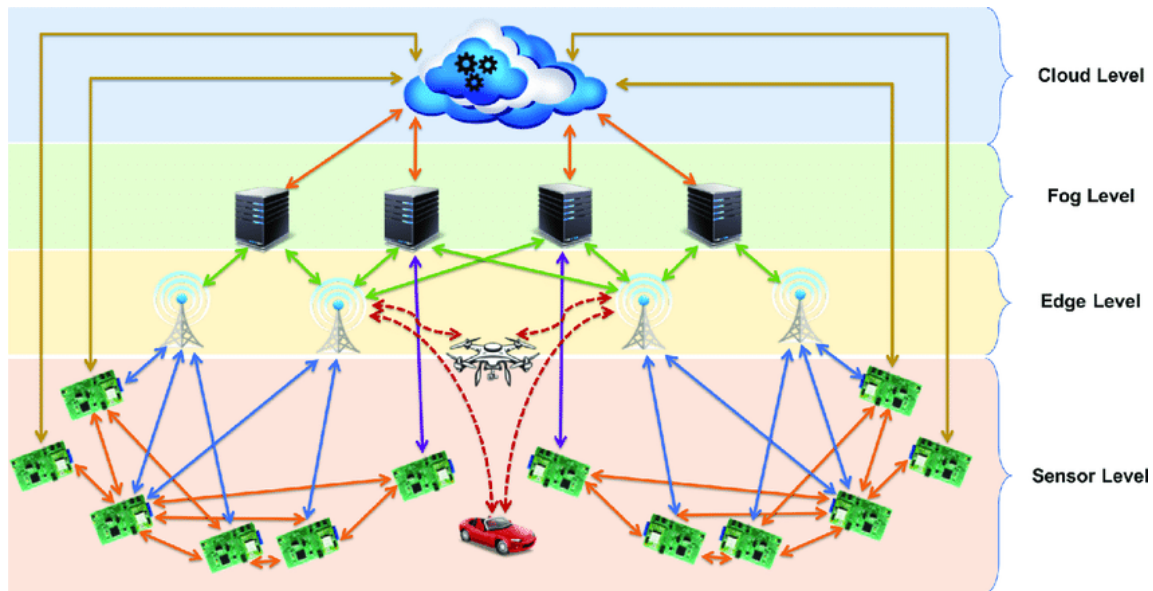
Pour répondre à ces problématiques, l'edge computing a vu le jour. C'est une méthode d'optimisation employée dans le cloud computing consistant à traiter les données à la périphérie du réseau, près de la source des données. Cette technique est apparue pour résoudre les faiblesses du cloud comme le problème de latence, permet de diminuer les coûts et d'apporter de la puissance des objets connectés. Prenons l'exemple d'une caméra de surveillance qui enverrait en temps réel les images qu'elle collecte, ces images sont ensuite traitées par un réseau de neurones permettant d'effectuer une reconnaissance faciale. Le principal coût du dispositif est lié à la quantité de données à envoyer à travers le réseau et traiter via le cloud.

Le principe, comme on peut le voir sur l'image<sup>4</sup> ci-dessous, consiste à traiter les données le plus proche de leur source le plus possible via l'edge computing. Les données ne pouvant pas être traitées via l'edge devront l'être par le fog computing. Les traitements ne pouvant être effectués ni par l'edge ni par le fog seront traités sur le cloud. Ce principe a pour but de limiter les données transitant sur le réseau pour améliorer la vitesse des traitements.

<sup>2</sup> <https://www.thirdlawreaction.com/tesla-drone-hack-explained/>

<sup>3</sup> <https://www.bleepingcomputer.com/news/technology/ovh-data-center-burns-down-knocking-major-sites-offline/>

<sup>4</sup> [https://www.researchgate.net/figure/loT-augmented-with-Fog-edge-Computing-Architecture\\_fig1\\_342083840](https://www.researchgate.net/figure/loT-augmented-with-Fog-edge-Computing-Architecture_fig1_342083840)  
[accessed 17 Oct, 2021]



### 3. TinyML ou machine learning embarquée : les microcontrôleurs sous stéroïdes ?

L'apparition de la discipline du TinyML vient pour répondre au besoin du Edge computing, principalement en proposant des applications pouvant être implémentées sur des environnements limités en ressources. Ainsi, plusieurs recherches ont été menées sur ce sujet, qui ont aboutit à des méthodes d'optimisation des modèles pour être exécutés sur des microcontrôleurs ou sur des accélérateurs d'inférence. Les géants de la technologie se sont vite tournés vers ce domaine à forte croissance. Ainsi nous avons vu le lancement de plusieurs gammes d'accélérateurs d'inférence ainsi que les frameworks de développement pour chaque plateforme.

La figure<sup>5</sup> suivante montre deux accélérateurs de la marque Intel (Neural Compute Stick 2)<sup>6</sup> fonctionnant sous OpenVino et celui de Google (Google Coral)<sup>7</sup> utilisant Tensorflow lite ou Tensorflow Micro.

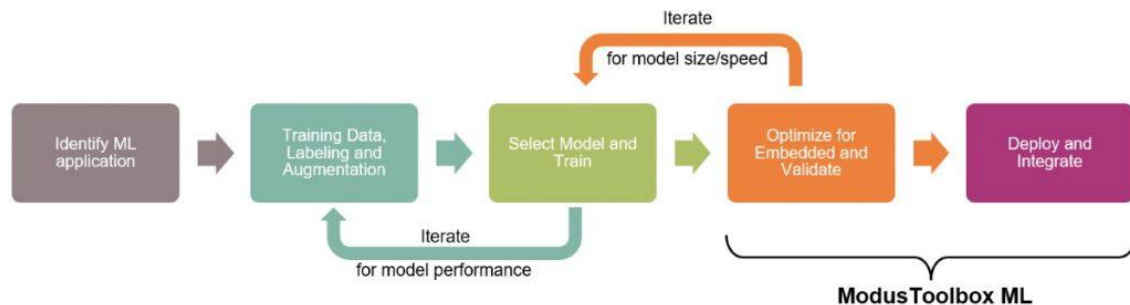
<sup>5</sup> [https://www.rs-online.com/designspark/rel-assets/dsauto/temp/uploaded/11\\_Movidius\\_Coral2.png?w=1042](https://www.rs-online.com/designspark/rel-assets/dsauto/temp/uploaded/11_Movidius_Coral2.png?w=1042)

<sup>6</sup> <https://www.intel.com/content/www/us/en/developer/tools/neural-compute-stick/overview.html>

<sup>7</sup> <https://www.coral.ai/>

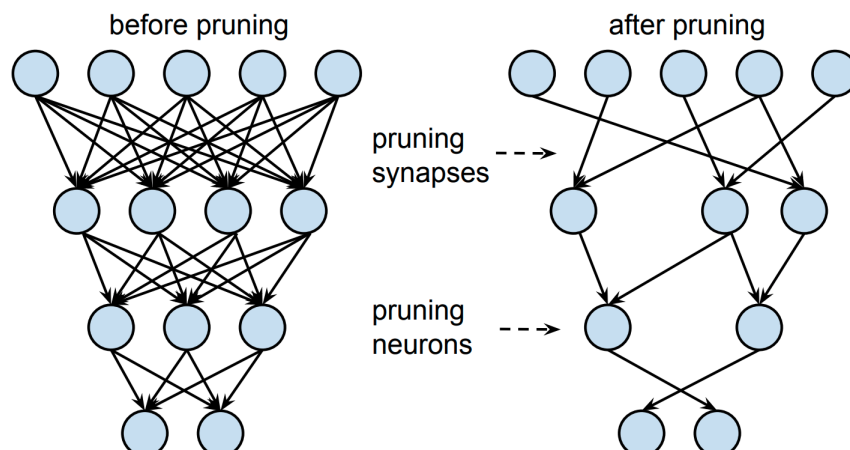


Le développement des modèles spécifiques à ces environnements hardware suit un cheminement bien particulier, car il doit prendre en compte les spécificités de chaque composant hardware, ce workflow est présenté dans la figure<sup>8</sup> suivante:



On constate qu'au plus des étapes d'entraînement connus du modèle, s'ajoute une étape d'optimisation qui est spécifique à la plateforme ciblée. Généralement les frameworks de développement tels que Tensorflow ou PyTorch proposent des méthodes qui permettent de proposer une version plus optimisée. Les méthodes d'optimisation les plus courantes sont :

- Le pruning, ou l'élagage des réseaux de neurones<sup>9</sup> : qui consiste à désactiver ou d'enlever des neurones sans avoir une perte considérable de précision. ce processus est montré sur la figure suivante:



<sup>8</sup> <https://www.eejournal.com/article/modus-toolbox-ml-tinyml-and-the-aiot/>

<sup>9</sup> <https://towardsdatascience.com/pruning-neural-networks-1bb3ab5791f9>



- La “quantization” des poids <sup>10</sup>: Ce processus consiste à réduire l’effet mémoire des poids qui sont représentés pendant l’entraînement en virgule flottante, le but étant d’avoir la représentation la plus légère en mémoire sans perte de précision.

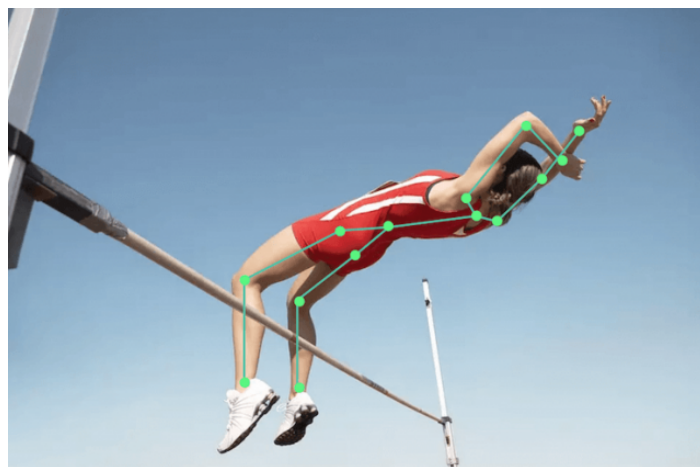
## Méthodes d’analyse et reconnaissance de gestes :

Parmi les domaines qui suscitent de plus en plus d’attention, notamment dans le domaine médical ou sportif ou encore dans le domaine des ventes. Par exemple, ces technologies permettent d’avoir une idée plus claire sur les comportements des clients dans un centre commercial, ce qui permet d’adapter l’emplacement des produits pour éviter aux clients de se mettre à genoux par exemple pour récupérer un produit en bas du rayon.

### 1. Analyse des flux vidéo/images :

L’approche la plus utilisée pour remédier à ces besoins est basée sur la vision par ordinateur (computer vision), qui permet d’avoir une estimation des gestes effectués par une ou plusieurs personnes en analysant une image ou une vidéo provenant d’une caméra fixée dans un périmètre précis. Cette analyse est effectuée à l’aide d’une parmi plusieurs architectures<sup>11</sup> à base de réseaux de neurones convolutifs.

Les avantages de cette approche est un gain considérable de précision par rapport aux autres méthodes, vu qu’elle exploite des images regroupant un ensemble de données spatiales qui sont perdues dans d’autres méthodes. L’un des points faibles de cette méthode étant le manque de portabilité lié à l’usage d’une caméra. Un exemple d’usage est présenté sur la figure<sup>12</sup> suivante:



<sup>10</sup> M. Wess, S. M. P. Dinakarrao and A. Jantsch, "Weighted Quantization-Regularization in DNNs for Weight Memory Minimization Toward HW Implementation,".

<sup>11</sup> <https://itrexgroup.com/blog/human-pose-estimation-use-cases-implementation-tips/#>

<sup>12</sup> <https://indatalabs.com/blog/3d-pose-estimation>

## 2. Vers des systèmes plus libres : usage de capteurs.

Une autre approche aux problèmes de reconnaissance de gestes étant celle qui exploite des données de capteurs. Les capteurs les plus utilisés sont des accéléromètres et des gyroscopes. Ces derniers offrent un avantage de mobilité intéressant pour des applications tels que des montres connectés ou des bracelets de fitness. Un autre avantage étant leurs coûts très compétitifs par rapport aux caméras. De plus, ces capteurs sont généralement de très petite taille, offrant la possibilité d'avoir des produits de petite taille, ce qui est pratique vu que le produit sera porté par les utilisateurs.

### **Cas d'utilisation et exemples d'usage éventuel :**

- Synchronisation des feux de signalisation avec les gestes des policiers de route:

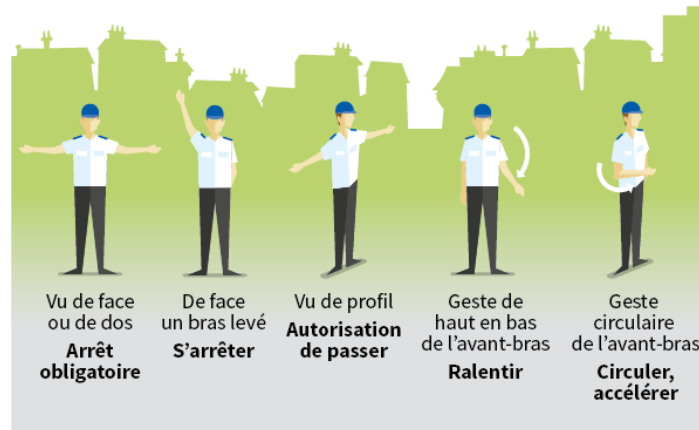
La gestion du trafic de véhicules en ville est un problème de plus en plus pressant, certaines villes ont adopté des moyens de gestion intelligents tels que des feux dont la durée est calculée à partir des flux de véhicules. Cependant, la présence des policiers est des fois indispensable dans les cas d'accidents ou d'embouteillages afin de fluidifier la circulation. Or, cette présence peut elle-même causer des problèmes liés au manque de visibilité des policiers pour les conducteurs. Ainsi, une confusion peut s'installer auprès des conducteurs lorsqu'ils s'aperçoivent que le feu est vert, alors que tous les véhicules sont à l'arrêt, certains conducteurs se mettent à klaxonner, d'autres essaieront même des manoeuvres qui peuvent mettre en danger la vie des piétons et des conducteurs à côté.

Ce qui explique le besoin de synchroniser les feux de signalisation avec les mouvements du policier. Cette synchronisation se base sur une reconnaissance de gestuelle de ce dernier qui doit être effectuée en temps réel, et doit avoir une précision élevée pour ne pas être une autre cause d'accidents. L'usage d'une reconnaissance à base d'une caméra n'est pas adapté à ce contexte, malgré le fait qu'il donnera plus de précision. L'usage des accéléromètres est un choix optimal pour cette problématique, car il donne plus de liberté de mouvement au policier, et lui permet d'utiliser cette fonction dans des carrefours non équipés de dispositifs de surveillance.

Un exemple des gestes ciblés par ce cas d'usage est donné sur la figure<sup>13</sup> suivante :

---

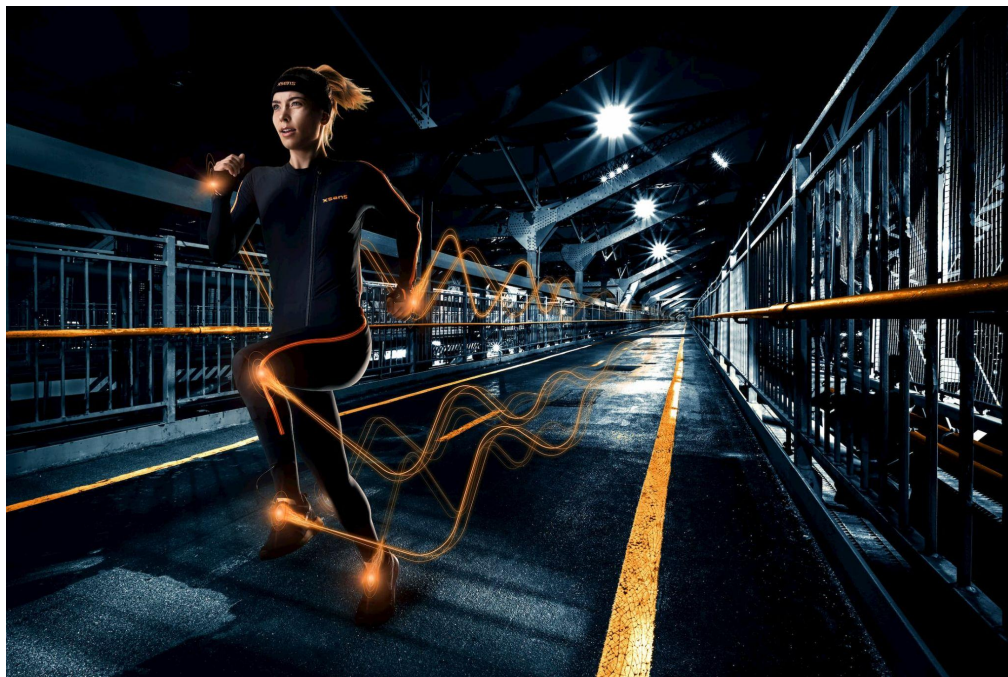
<sup>13</sup> <https://www.ornikar.com/code/cours/circulation/intersections>



- Etude des mouvements des sportifs professionnels:

Un autre cas d'usage étant l'étude en vue d'amélioration des mouvements des sportifs de haut niveau. Dans des sports spécifiques, par exemple les sports de lancer, le mouvement des bras ainsi que ceux du corps en entier jouent un rôle crucial dans la vitesse initiale de lancée qui impacte la distance parcourue par l'objet lancé. Ainsi, l'étude des mouvements de ces athlètes leur permet d'optimiser leurs mouvements. Une approche peut alors être la quantification des mouvements à l'aide d'un ensemble de capteurs, afin de visualiser et étudier chaque étape de leurs mouvements de lancer.

Cette approche peut être étendue à d'autres sports, toujours dans le même objectif, un exemple d'usage est présenté sur la figure<sup>14</sup> suivante :

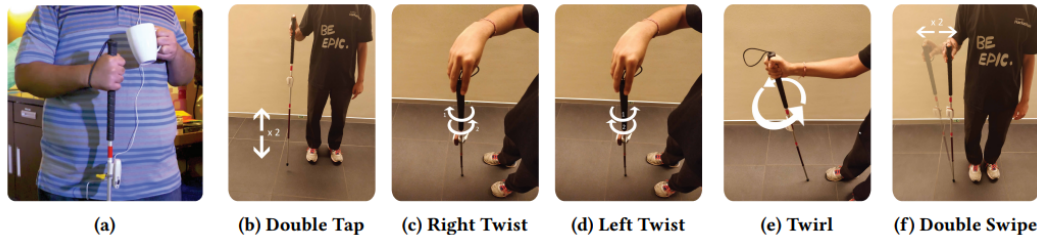


<sup>14</sup><https://www.xsens.com/press-releases/demonstration-worlds-first-wearable-3d-body-motion-tracking-system-based-consumer-grade-mems-sensors-international-cs>



- Ajout de fonctionnalités sur des cannes destinés aux malvoyants :<sup>15</sup>

La commercialisation des cannes connectées à destination des malvoyants a permis une augmentation des innovations dans le domaine d'accessibilité. Ainsi, des fonctionnalités de détection d'obstacles, d'assistance vocale ont vite été intégrées dans des produits similaires. Une extension possible est d'utiliser la reconnaissance des gestes effectués avec la canne pour activer des fonctionnalités supplémentaires. L'activation de ces fonctionnalités à l'aide de mouvements au lieu des boutons permet de réduire l'encombrement de la canne, ce qui rend son usage plus simple. De plus, cette méthode permet d'ajouter des fonctionnalités (des gestes) sans avoir à refaire le design de la canne, ce qui permet de réduire les coûts de production.



## Limitations et problématiques :

Malgré la multitude des cas d'usage possibles de la reconnaissance des gestes à partir des accéléromètres, plusieurs limitations ont été recensés dans la littérature scientifique, on peut en citer les limitations suivantes :

- Problèmes liés à l'orientation des capteurs<sup>16</sup> : L'orientation ainsi que la position initiale de l'accéléromètre peuvent causer de grandes variations sur le signal. Ainsi, on peut avoir de mauvais résultats qui sont liés à la manière d'usage ou l'endroit de placement du capteur. d'où la nécessité d'une séquence de calibration du capteur avant la mise en action.
- Problèmes liés à la variation des vitesses de mouvement : Il est clair qu'un enfant n'a pas la même vitesse de mouvement qu'une personne âgée, d'où l'intérêt de choisir une fenêtre assez large pour capturer tout le signal dans le cas de faibles vitesses de mouvement, ainsi que l'enrichissement de la dataset par des exemples de mouvements avec des vitesses variés

<sup>15</sup> GesturePod: Gesture-based Interaction Cane for People with Visual Impairments Shishir G. Patil Don K. Dennis Chirag Pabbaraju Nadeem Shaheer Harsha Simhadri Vivek Seshadri Manik Varma Prateek Jain.

<sup>16</sup> Chavarriaga, Ricardo & Bayati, Hamidreza & Millan, Jose del R.. (2013). Unsupervised adaptation for acceleration-based activity recognition: Robustness to sensor displacement and rotation. Personal and Ubiquitous Computing. 17. 10.1007/s00779-011-0493-y.