Detección de trastornos del espectro autista por video

Autora:
Anna Arias Duart

Tutores:
Séverine Dubuisson
Djamal Merad

Escuelas:
Télécom ParisTech
Université Pierre et Marie Curie

Titulación:
Master Informatique

Laboratorio:
Laboratoire d'Informatique & Systèmes

Curso:
2017-2018

Résumé

La détection du Trouble du Spectre Autistique (TSA) repose sur l'observation intensive de vidéos afin d'analyser le comportement naturel de l'enfant et de déceler d'éventuels troubles caractérisés par des anormalités dans les interactions sociales, la communication ainsi que par un comportement spécifique. La vision par ordinateur constitue donc une solution efficace pour étudier le comportement des enfants autistes.

Dans ce travail, nous présentons une méthode non invasive de détection et reconnaissance de quatre mouvements stéréotypés chez les enfants autistes. Ces mouvements sont les suivants : le battement des mains, les mains dans le visage, les mains dans les oreilles et les coups de tête ou balancement du corps.

Pour réaliser cette tâche, nous utilisons une modification de l'algorithme *Dynamic Time Warping* (DTW) classique afin d'aligner les différentes trajectoires des angles formés par les paires d'articulations de nos corps. Notez que les coordonnées 3D des articulations sont obtenues par un capteur 3D de type Kinect.

Abstract

The Autism Spectrum Disorder (ASD) detection is based on the intensive observation of videos to analyze the child's natural behavior and to detect possible disorders characterized by anomalies in social interactions, communication as well as specific behavior. Therefore, computer vision is an effective solution to study the behavior of autistic children.

In this work, we present a non-invasive method for detecting and recognizing four stereotyped movements of autistic children. These movements are : flapping hands, hands on the face, tapping ears and head banging.

To accomplish this task, we use a modification of the Dynamic Time Warping (DTW) algorithm in order to align the different trajectories of the angles formed by the pairs of joints of our bodies. Note that the 3D joint coordinates are obtained by a 3D Kinect v2 sensor.

Resumen

La detección del Trastorno del Espectro Autista (TEA) se basa, en muchos casos, en la observación intensiva de videos para analizar el comportamiento natural del niño o niña y detectar posibles trastornos caracterizados por anomalías en las interacciones sociales, la comunicación así como por comportamientos específicos. Por este motivo, la visión por ordenador es una solución eficaz para estudiar el comportamiento de los niños y niñas autistas.

En este trabajo, se presenta un método no invasivo de detección y reconocimiento de cuatro movimientos estereotipados en niños y niñas con autismo. Estos movimientos son los siguientes : el aleteo de manos, las manos en la cara, las manos en las orejas y el balanceo del cuerpo o golpes de cabeza.

Para realizar esta tarea, se ha utilizado una modificación del algoritmo *Dynamic Time Warping* (DTW) clásico con la finalidad de alinear las diferentes trayectorias de los ángulos formados por los pares de articulaciones del cuerpo humano. Se debe remarcar que las coordenadas 3D de las articulaciones son obtenidas por un sensor 3D de tipo Kinect.

Table des matières

1

Introduction

Selon la Bibliothèque Nationale de Médecine des États-Unis le Trouble du Spectre Autistique (TSA) est "un trouble neurologique et développemental qui commence tôt dans l'enfance et dure toute la vie d'une personne. Il affecte la façon dont une personne agit et interagit avec les autres, communique et apprend".

La détection du TSA est basée sur des observations périodiques, réalisés par des professionnels, des comportements spécifiques des enfants. Le diagnostic du TSA peut être considéré comme fiable à l'âge de 2 ans, cependant, de nombreux enfants ne reçoivent pas le diagnostic définitif tant qu'ils ne sont pas beaucoup plus grands.

Comme Hashemi *et al.* l'expliquent dans [6] différentes études ont été menées afin de fournir des outils de dépistage quantitatifs et automatiques qui aident les médecins à diagnostiquer les enfants à risque de TSA le plus tôt possible.

Bien que le TSA n'a pas de remède, une détection précoce leur permettra de recevoir l'aide dont ils ont besoin augmentant ainsi, les chances d'amélioration chez l'enfant.

Ce document est organisé comme suit. Dans le deuxième chapitre d'abord nous décrivons brièvement comment se caractérise l'autisme dans le comportement et nous donnons quelques exemples des différents comportements dans lesquels différents auteurs se sont focalisés pour mener à bien leur travail. Ensuite, nous présentons les deux grands groupes dans lesquels nous pouvons diviser les différentes méthodes proposées pour la détection de l'autisme : les techniques invasives et les techniques non invasives.

Le troisième chapitre est consacré à la description du capteur utilisé ainsi qu'à la discussion de ses limites.

Le quatrième chapitre peut être divisé en trois parties. Dans la première partie, nous étudions l'état de l'art de la reconnaissance des actions humaines

et présentons trois méthodes qui permettent de trouver la correspondance entre un signal et un modèle. Ensuite, dans la deuxième partie, nous décrivons la procédure qui nous a permis d'atteindre l'approche finale. Dans la troisième partie, les différents tests effectués sont montrés.

Ce travail se termine par une conclusion générale avec des perspectives.

2

Autisme et vision par ordinateur

D'après le *Manuel diagnostique et statistique des troubles mentaux* [1], les personnes atteintes de TSA ont en particulier :

- Des difficultés avec la communication et l'interaction avec d'autres personnes. Par exemple : une absence de contact visuel, ou une tendance à ne pas regarder ou écouter les gens ou bien des expressions ou mouvements qui ne correspondent pas à ce qui est dit.
- Des intérêts restreints et des comportements répétitifs. Par exemple : répéter des mots ou des phrases ou bien des intérêts trop concentrés, comme avec des objets en mouvement.
- Les symptômes qui affectent le développement de la vie d'une personne : au travail, à l'école ou dans d'autres domaines. Par exemple, ils ont des problèmes d'empathie avec les autres.

Différentes recherches ont été faites dans ce domaine mais comme l'autisme se caractérise de différentes manières et il n'y a pas une seule caractéristique typique ou un seul comportement discriminant pour détecter l'autisme, les approches ont focalisé sur des comportements différents, comme nous allons en donner quelques exemples ci-dessous.

Hashemi *et al.* proposent dans [6] des outils de vision artificielle pour observer quatre modèles comportementaux spécifiques liés aux TSA : le suivi visuel, le désengagement de l'attention, le partage d'intérêt et le comportement moteur atypique (concrètement les auteurs se centrent sur le fait que les enfants diagnostiqués présentent souvent des positions de bras asymétriques).

D'autres se sont concentrés sur des mouvements stéréotypés. Par exemple, dans [4] les auteurs ne détectent qu'un seul mouvement : battement des mains. Dans [7] les auteurs détectent quant à eux 5 gestes : le balancement du corps, le battement de mains, le battement des doigts, la main dans le visage et la main derrière le dos. Des travaux ont travaillé sur jusqu'à 10 mouvements [10], incluant également des comportements dynamiques tels

que la marche en cercles.

Dans le cadre de ce stage nous nous sommes focalisé sur quatre comportements : le battement des mains, les mains dans le visage, les mains dans les oreilles et le balancement du corps ou coups de tête (*head banging*).

2.1 Techniques invasives et non invasives

Parmi les différentes approches mises en œuvre, les approches invasives peuvent être différenciées des approches non invasives. Dans les approches dites invasives, le sujet doit porter un dispositif, par exemple un bracelet. En revanche, avec les approches dites non invasives le sujet est libre et lors de la détection son environnement n'est pas perturbé.

- Un exemple d'approche **invasive** est celui proposé dans [5] où les auteurs utilisent des accéléromètres selon 3 axes pour détecter des mouvements stéréotypés comme le balancement du corps ou encore le battement des mains.
- Dans [6] nous trouvons un exemple d'approche **non invasive** où le sujet est libre et les auteurs ont tout simplement enregistré les comportements des enfants afin d'observer les 4 modèles comportementaux que nous avons déjà cités : le suivi visuel, le désengagement de l'attention, le partage d'intérêt et le comportement moteur atypique. Ainsi dans [7] ou [10], dont les travaux ont été plus détaillés ci-dessus, nous trouvons également des approches non invasives où les auteurs détectent de mouvements stéréotypés à l'aide du capteur Kinect.

Pour mettre en évidence les avantages et les inconvénients de chaque type d'approche, nous nous référerons à l'article [4] où les auteurs proposent une approche non invasive et une approche invasive afin de détecter le mouvement stéréotypé du battement des mains.

Dans l'approche non invasive les auteurs utilisent le capteur Microsoft Kinect. Et dans l'approche invasive les auteurs utilisent un dispositif de Texas Instruments avec des accéléromètres intégrés, la montre eZ430-Chronos.

Les résultats que les auteurs ont obtenus sont les suivants : avec la Kinect, les auteurs ont détecté 51% des mouvements stéréotypés, tandis qu'avec la montre eZ430-Chronos électronique il en ont obtenu 76%. Il faut remarquer que 49% des faux positifs obtenus avec la Kinect correspondaient à l'intervention d'un chercheur : les travaux reposant sur la Kinect ne distinguaient pas l'enfant du chercheur.

Les résultats obtenus avec le eZ430-Chronos sont plus précis, mais ont l'inconvénient devoir porter un capteur au poignet. Cette situation peut être inconfortable pour certains enfants (l'un des enfants de leurs sessions a tenté

d'enlever le bracelet).

Dans le cadre de notre projet, nous avons décidé d'utiliser des approches non invasives qui nous semblent plus adaptées pour travailler avec des enfants autistes. Nous allons donc utiliser le capteur Kinect pour enregistrer les différents comportements. Ce capteur est décrit dans le chapitre qui suit.

3

Capteur

Afin de détecter les gestes ou mouvements stéréotypés chez les enfants autistes nous avons décidé d'utiliser la méthode de détection des points d'intérêt à l'aide du capteur Kinect v2 de Microsoft. Nous présentons brièvement ce capteur dans la section 3.1 et en donnons les limites observées à travers nos tests dans la section 3.2.

3.1 Présentation du capteur

La Kinect possède deux types de caméra : une caméra RVB et une caméra infrarouge (IR) pour détecter la profondeur. La Kinect dispose également d'un microphone *multi-array* qui permet de reconnaître la voix ainsi que de positionner sa provenance.

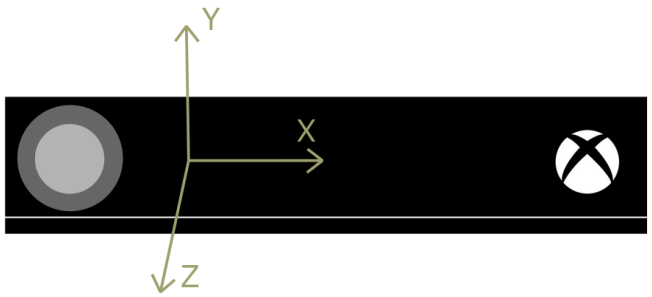


FIGURE 3.1 – Kinect v2 et système de coordonnées du capteur.

La Kinect v2 est capable de reconnaître jusqu'à 25 points du corps humain de jusqu'à 6 personnes simultanément, voir la figure 3.2.

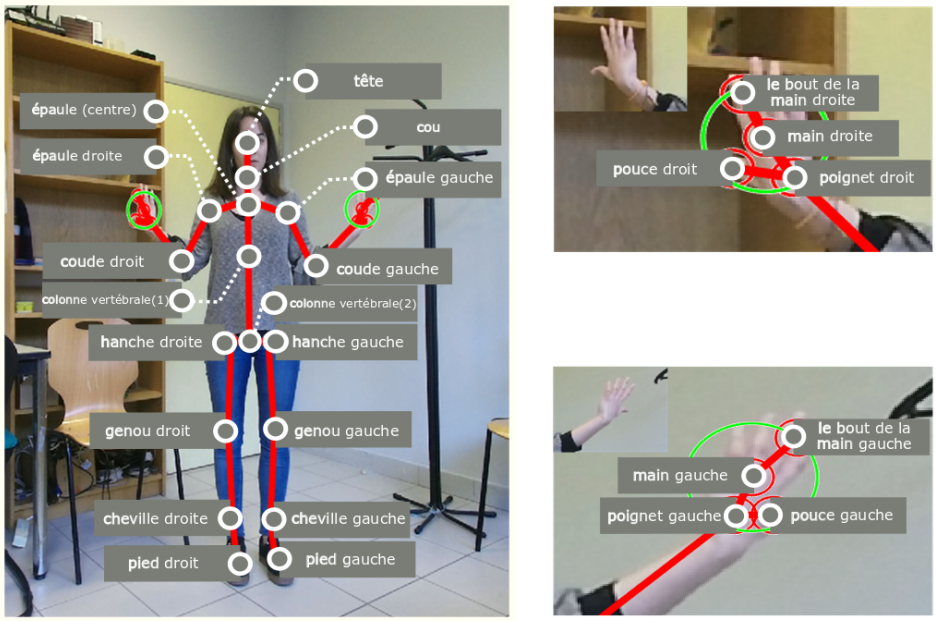


FIGURE 3.2 – Les 25 points suivis par le capteur Kinect v2.

Pour chaque articulation nous avons trois dimensions, ce qui permet de positionner chaque articulation par rapport à la Kinect. C'est-à-dire étant située l'origine du capteur IR sur la Kinect, voir figure 3.1 : l'axe des X va vers la gauche du capteur, celui des Y va vers le haut (cette direction depend de l'inclinaison du capteur) et celui des Z va dans la direction du capteur (en d'autres termes Z représente la distance du capteur).

3.2 Tests et limites

Nous avons effectué plusieurs tests pour pouvoir étudier les limites de la Kinect.

— Enregistrement
Afin d'analyser les vidéos, la première étape est de pouvoir les stocker efficacement. Cette étape est d'une importance vitale car si par exemple nous ne stockons que quelques images par seconde, nous pouvons perdre des mouvements de patients qui peuvent être spécifiques au trouble qui nous intéresse.
Notre cerveau peut traiter de 10 à 12 images par seconde, c'est-à-

dire que nous pouvons percevoir individuellement. En revanche, si le nombre d'images par seconde est supérieur à 12 nous allons percevoir un mouvement. Nous avons considéré donc nécessaire pouvoir stocker plus de 12 images par seconde.
Finalement, nous avons réussi à stocker 400 images dans un temps moyen d'environ 20 secondes. Ce qui équivaut à environ 20 images par seconde, un résultat qui semble satisfaisant pour pouvoir analyser correctement les mouvements.

— Profondeur maximale et suivie du squelette
Après avoir effectué différents tests, nous pouvons conclure que, comme les spécifications marquent, la profondeur maximale que la Kinect peut détecter est d'environ 8 mètres, voire même un peu moins.
Par contre, ce qui nous intéresse pour la réalisation de nos mesures est la distance maximale à laquelle la Kinect est capable de détecter correctement le squelette.
Après avoir enregistré différentes séquences à l'intérieur nous pouvons donc conclure que bien que la profondeur maximale que la Kinect puisse détecter soit de 8m, la distance maximale à laquelle la Kinect est capable de suivre correctement le squelette est à une distance d'environ 4.5m.
Ces spécifications correspondent aux mesures effectuées à l'intérieur. Nous devons souligner que la Kinect est moins précise à l'extérieur. Il faut se rappeler que la Kinect utilise une caméra infrarouge pour détecter la profondeur : plus la lumière du soleil est directe, plus il y aura d'interférence et donc la précision de la profondeur sera pire. Dans tous les cas, toutes les expériences que nous avons faite par la suite ont été menées à l'intérieur.

— Champ de vision
Selon les spécifications, le champ de vision horizontal est de 70° et le champ de vision vertical est de 60°. Si nous ajoutons à cela que la distance maximale à laquelle le squelette peut-être détecté est d'environ 4.5 mètres et prenons par exemple un sujet de 1.5 mètre, nous obtenons que nous ne pouvons nous déplacer que d'environ 3.2 mètres de profondeur et en tenant en compte du fait que si nous nous approchons à la Kinect le champ de vision horizontal diminue aussi, voir figure 3.3.

La réalisation de ces tests nous a permis de définir le champ de vision de la Kinect mais également les conditions dans lesquelles les acquisitions vont pouvoir se faire. Ces séquences d'images acquises seront utilisées ci-dessous pour la détection des quatre mouvements stéréotypés. La procédure de reconnaissance de ceux-ci est expliquée dans le chapitre 4.

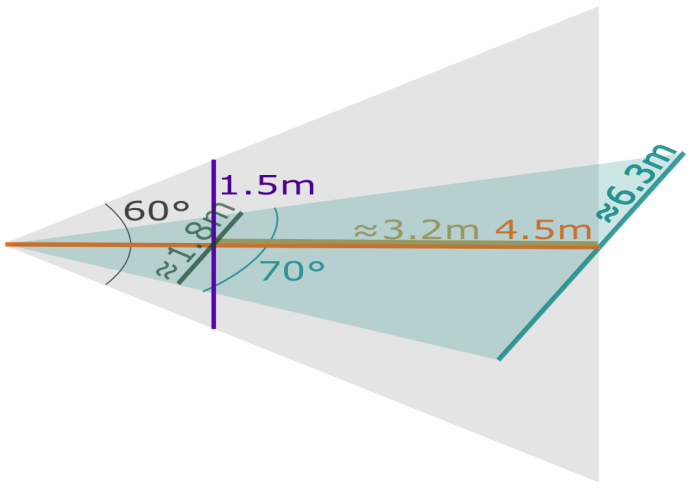


FIGURE 3.3 – Champ de vision de la Kinect.

4

Mouvements stéréotypés

Comme nous l'avons vu dans le chapitre 2 l'un des symptômes des enfants autistes est l'exécution de mouvements ou gestes spécifiques. Pour les reconnaître, l'idée est de trouver une correspondance entre le mouvement courant et un modèle qui a été enregistré précédemment.

Différentes approches ont mis l'accent sur la reconnaissance des actions humaines. Certaines utilisent des algorithmes de classification, comme l'approche proposée par Chenyang *et al.* [2] qui utilise des angles entre les articulations et la distance entre la tête et le plan du sol comme caractéristiques pour entraîner un classifieur SVM (*Support Vector Machine*). D'autres utilisent des réseaux de neurones convolutionnels (CNNs). Par exemple Laraba *et al.* [8] représentent une séquence des articulations du squelette 3D dans un espace 2D (espace RVB). Ainsi la classification des actions devient un problème de classification des images, permettant ainsi d'utiliser les modèles CNN existants.

Pour ces méthodes de classification nous devons avoir de grandes bases de données et ceci est difficile pour le cas qui nous concerne, c'est-à-dire pour les enfants autistes pour lesquelles nous n'avons pas suffisamment de données. C'est pourquoi nous avons choisi de ne pas utiliser des méthodes de classification et de nous focaliser sur des méthodes qui cherchent la mise en correspondance avec un mouvement modèle, sans aucun type d'apprentissage et donc sans avoir besoin d'une grande base de données.

4.1 Travaux existants

Différentes méthodes existent pour faire cette correspondance, nous allons détailler trois d'entre elles.

4.1.1 *$ Recognizer*

Cette approche a d'abord été pensée pour la reconnaissance gestuelle. L'interaction gestuelle nécessite des approches de reconnaissance gestuelle rapides, simples et précises. La famille $ de reconnaisseurs ($1, $N) répond à ce besoin : $1 traite les gestes continus, en revanche les approches $N peuvent traiter des gestes non continus, en reliant parties continues.

Cependant, comme l'ordre et la direction des traits peuvent différer entre les utilisateurs dessinant le même symbole, $N doit générer toutes les permutations possibles, ce qui provoque une explosion à la fois en mémoire et en temps d'exécution.

Vatavu *et al.* proposent dans [14] une méthode qui a été appelée $P *Recognizer*, cette approche évite la complexité de stockage de $N en représentant les gestes comme des "nuages de point" et ignore ainsi le comportement variable de l'utilisateur en termes d'ordre et de direction des traits.

L'idée est de trouver pour chaque point du premier nuage (C_i), le point le plus proche du second nuage qui n'a pas encore été apparié. Une fois que le point C_i est apparié, on considère C_{i+1} jusqu'à ce que tous les points de C soient appariés ($i = 1 \dots n-1$).

Exemple d'application : Jazouli *et. al* proposent dans [7] une méthode pour la détection automatique de 5 mouvements stéréotypés comme nous l'avons expliqué dans le chapitre 2. Cette approche utilise la méthode du $P *Recognizer* pour la reconnaissance de gestes considérés comme un nuage de points.

Afin de pouvoir comparer un nouveau geste avec le geste de référence (modèle) il faut effectuer une normalisation : tout d'abord, il faut rééchantillonner le nouveau geste, ensuite, le geste est mis à l'échelle par rapport au rectangle de référence et translaté de façon que le centroïde soit en $(0, 0, 0)$.

Finalement, la liste des points obtenus est comparée à chacun des modèles existants, afin de trouver ou pas une correspondance avec un des mouvements stéréotypés cherchés.

4.1.2 Correspondance de courbes

Dans [3], Cui *et al.* présentent une nouvelle signature de courbe qui est invariante à l'échelle et peut alors traiter des transformations de similarité.

La méthode peut être divisée en deux parties, la première est l'extraction de la signature (invariante à l'échelle) et la deuxième consiste à utiliser la signature pour faire l'appariement.

1. **Extraction de la signature :** la première étape consiste à convertir la courbe d'entrée en une représentation paramétrique. Ensuite, il faut échantillonner la courbure de la courbe à des intervalles de longueur d'arc égaux. L'étape suivante consiste à intégrer les courbures non signées discrètement le long de la courbe en additionnant les valeurs absolues des courbures discrètement échantillonnées le long de la courbe. Et finalement, on calcule la courbure à des points échantillonnés à intervalles égaux le long de l'intégrale de l'axe de courbure non signée. Cette courbe sera notre signature qui est invariante à l'échelle. Autrement dit, les parties de la signature potentiellement appariées couvriront la même longueur même si les courbes d'entrée ont une différence d'échelle.
2. **Correspondance de courbe :** pour le cas qui nous concerne, c'est-à-dire l'appariement d'une courbe entière à une partie d'une autre ; les auteurs utilisent la corrélation croisée normalisée pour évaluer la similitude entre ces deux courbes.

Exemple d'application : Cui *et al.* présentent dans [3] différentes applications de cet algorithme, comme par exemple des exemples d'appariement des traits manuscrits, en obtenant des résultats satisfaisants.

Toutefois, le principal inconvénient de cet algorithme est qu'il est conçu pour faire correspondre deux courbes sous une transformation de similarité, c'est-à-dire sous une translation, rotation ou bien sous un changement d'échelle uniforme.

4.1.3 *Dynamic Time Warping*

Cette approche est un algorithme de normalisation temporelle pour la reconnaissance de la parole introduit dans l'article [12] en 1978.

La variation de la fréquence quand nous parlons provoque une fluctuation non linéaire dans l'axe temporel, le but de cet algorithme est de trouver la meilleure correspondance entre deux séries temporelles et ainsi d'essayer d'éliminer cette fluctuation.

Notez que cet algorithme sera expliqué plus en détail car il sera utilisé plus tard dans la section 4.2.

Les deux séries que nous voulons comparer X et Y sont définies comme suit :

$$\begin{aligned} X &:= (x_1, x_2, \ldots x_i, \ldots x_N) \\ Y &:= (y_1, y_2, \ldots y_j, \ldots y_M) \end{aligned} \tag{4.1}$$

Pour trouver le meilleur alignement entre X et Y, il faut trouver le chemin optimal à travers la matrice des coûts : $c \in \mathbb{R}^{N \times M}$. Le chemin optimal sera donc celui avec lequel nous obtenons le coût global minimum.

Matrice distance : pour obtenir la matrice des coûts, nous devons d'abord calculer la matrice distance. La fonction distance d peut être définie de différentes manières, nous utilisons la suivante :

$$d(i,j) = \sqrt{(x_i - y_j)^2} \quad où \quad i = 1...N \quad et \quad j = 1...M \tag{4.2}$$

Matrice de coûts : une fois calculée la matrice distance, nous procédons à calculer la matrice de coûts :

$$c(i,j) = \begin{cases} d(1,1) & \text{si } i = 1 \text{ et } j = 1 \\ c(i-1,1) + d(i,1) & \text{si } i \neq 1 \text{ et } j = 1 \\ c(1,j-1) + d(1,j) & \text{si } i = 1 \text{ et } j \neq 1 \\ min\{c(i,j-1), c(i-1,j-1), c(i-1,j)\} & \text{si } i > 1 \text{ et } j > 1 \end{cases} \tag{4.3}$$

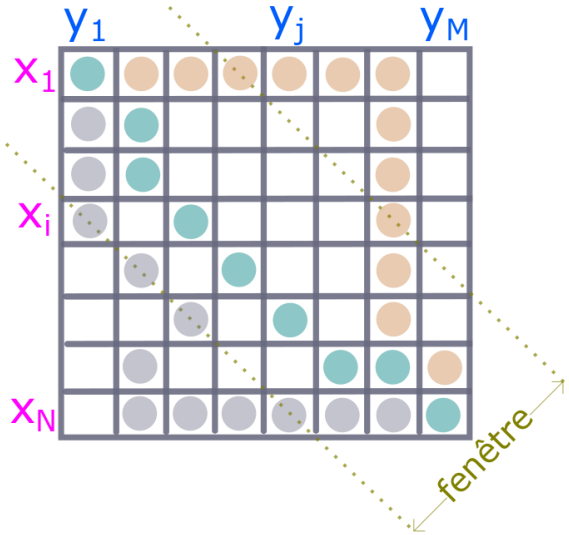


FIGURE 4.1 – Matrice de coûts. Plusieurs chemins à travers la grille sont possibles, afin de limiter la recherche et trouver la meilleure solution, nous définissons les conditions suivantes : condition de monotonie, condition de continuité, condition aux limites, condition de la fenêtre de réglage et la condition de la pente.

Chemin de déformation (*warping path*) : l'étape suivante sera de trouver le chemin optimal, avec lequel nous obtenons le coût le plus bas.

Nous définissons le chemin de déformation comme suit :

$$p = (p_1, ..., p_L) \text{ où } p_l = (i_l, j_l) \in [1 : N] \text{ x } [1 : M] \text{ pour } l \in [1 : L] \tag{4.4}$$

Plusieurs chemins à travers la grille sont possibles, voir figure 4.1, donc certaines restrictions doivent être appliquées à l'algorithme DTW pour limiter l'espace d'étude :

— Condition de monotonie. Le chemin d'alignement ne remonte pas dans l'index temporel. C'est-à-dire :

$$i_1 \leq i_2 \leq ... \leq i_N \text{ et } j_1 \leq j_2 \leq ... \leq j_M \tag{4.5}$$

Un exemple de chemin qui ne remplirait pas la condition de monotonie serait le chemin en gris de la figure 4.1.
Cette condition évite que les caractéristiques soient répétées dans l'alignement.

— Condition de continuité. Le chemin d'alignement ne saute pas dans l'indice temporel.

$$i_l - i_{l-1} \leq 1 \text{ et } j_l - j_{l-1} \leq 1 \tag{4.6}$$

Cette condition garantit que chaque caractéristique de la série est considérée.

— Condition aux limites. Le chemin d'alignement commence en haut à gauche et se termine en bas à droit, voir les chemins représentés sur la figure 4.1.

$$p_1 = (1, 1) \text{ et } p_L = (N, M) \tag{4.7}$$

Cette condition correspond au fait que la fluctuation de l'axe de temps n'entraîne jamais une différence de temps trop importante.

— Condition de la fenêtre de réglage. Une bonne trajectoire d'alignement est peu susceptible de dévier trop loin de la diagonale.

$$|i_l - j_l| \leq \text{w} \tag{4.8}$$

où w correspond à la taille de la fenêtre.
Si nous regardons les trois chemins affichés dans la figure 4.1, cette condition ne serait remplie que par le chemin bleu.
Cette condition évite d'ignorer différentes caractéristiques et empêche l'algorithme de rester bloqué sur une seule valeur.

— Condition de la pente. Le chemin d'alignement ne doit pas avoir beaucoup de marches sur l'axe i ou sur l'axe des j. Par exemple, cette condition ne sera pas satisfait par le chemin en orange de la figure 4.1. Cette condition empêche qu'une petite partie de la série numérique X corresponde à une grande partie de la série numérique Y ou vice-versa.

Coût du chemin de déformation : le coût du chemin de déformation (p) sera défini par l'équation suivante :

$$cout_p(X,Y) = \sum_{l=1}^{L} c(i_l, j_l) \tag{4.9}$$

Distance DTW : parmi tous les chemins de déformation possibles, l'optimum (p^*) est celui avec lequel on obtient le moindre coût. La distance DTW entre X et Y sera donc le coût du chemin optimal :

$$DTW(X,Y) = cout_{p^*}(X,Y) \tag{4.10}$$

Alors que le chemin de déformation optimal permet de trouver le meilleur alignement entre les deux séries (X et Y), la distance DTW nous permet également de mesurer le degré de similarité entre eux : plus la distance DTW est petite, plus les signaux seront similaires.

Exemple d'application :

— Dans [13], Sempena *et al.* classent les différentes activités en utilisant l'algorithme du plus proche voisin avec l'algorithme DTW. Pour décrire les mouvements humains les auteurs utilisent l'orientation de chaque articulation par rapport au repère du monde. Ce vecteur de caractéristiques est donc invariant à la rotation et à la taille du corps humain.
Les auteurs représentent les rotations sous la forme de quaternions parce que c'est plus compact, 4 nombres au lieu d'une matrice orthogonale de 9 chiffres pour chaque articulation (angles d'Euler) et en plus ils sont plus efficaces en termes de calcul par rapport aux angles d'Euler. Notez que les angles d'Euler sont les différentes rotations successives sur chaque axe (x, y, z) qui donnent naissance à la matrice de rotation orthogonale que nous connaissons.
Cette approche arrive à bien reconnaître des actions simples comme par exemple une vague avec des mains ou un applaudissement.

— Riofrío *et al.* proposent dans [11] une méthode pour la reconnaissance gestuelle des articulations supérieures du corps humain en utilisant la technique DTW avec le but de contrôler un diaporama.
Afin d'être invariant à la position du sujet, les auteurs proposent de déplacer l'origine du système de référence (dans le cas de la Kinect le système de référence est le centre de la Kinect) vers un point fixe par rapport au corps de l'utilisateur, les auteurs choisissent le centre des épaules.
Pour la reconnaissance de gestes, 10 mouvements différents sont acquis et enregistrés comme référence. Ensuite, ils comparent le *buffer* "test" afin de reconnaître s'il correspond à un des dix mouvements en utilisant l'algorithme DTW.
Cette approche présente aussi la possibilité de manipuler les paramètres internes du DTW en cherchant la meilleure configuration. Ces paramètres sont le seuil DTW, la condition de pente et la condition de la fenêtre de réglage.
— Rihawi *et al.* [10] fournissent une base de données en ligne de vidéos de profondeur de 10 modèles de comportements autistes. Les auteurs divisent ces comportements répétitifs en deux catégories principales :
  — Comportements statiques où il n'y a pas de mouvement réel au milieu de l'action.
  — Comportements dynamiques qui peuvent être faits avec certains mouvements pendant une période de temps.

  Rihawi *et al.* proposent aussi un algorithme de détection et utilisent la méthode DTW. Pour chaque trame les auteurs extraient tout d'abord les caractéristiques de toutes les articulations du squelette. Chaque articulation fournit deux caractéristiques θ_1 et θ_2 qui sont les angles euclidiens dans l'espace $3D$ entre YZ et ZX pour chaque trame. Ensuite, s'il n'y a pas de changement dans les valeurs des caractéristiques à partir d'un temps ϵ, ils essayent de détecter les comportements statiques. Sinon ils essayent de détecter une action répétitive en comparant les distances DTW du temps courant au temps précédent.
  Avec cet algorithme les auteurs réussissent à bien détecter des comportements statiques, en revanche les résultats de la détection des comportements dynamiques sont moins bons. Il faut remarquer que dans des actions, comme par exemple marcher en cercles, de nombreuses parties du corps du squelette pourraient être occultées et donc difficiles à détecter avec une Kinect.

4.2 Approches proposées

Dans la section précédente, nous avons vu différentes approches qui avaient comme objectif la comparaison des courbes dans le but de détecter si un modèle recherché avait eu lieu.

Notre objectif est donc de pouvoir détecter si un sujet a effectué l'un de ces quatre mouvements stéréotypés : couvrir le visage avec les mains, les mains dans les oreilles, le battement des mains ou les coups de tête (*head banging*). Pour le détecter, nous allons analyser la séquence d'images que nous avons stockée d'un nouveau sujet pour trouver le motif qui correspond au mouvement cherché, ce motif aura été précédemment enregistré.

Dans la section 4.2.1, nous présentons les caractéristiques utilisées par la suite. Dans la section 4.2.2 et dans la section 4.2.3 nous introduisons brièvement une première et une seconde approche : ces méthodes ainsi que leurs limites nous ont conduit à l'approche finale expliquée dans la section 4.2.4. Les résultats de cette dernière approche seront présentés et discutés dans la section suivante 4.3.

4.2.1 Extraction de caractéristiques

Comme nous l'avons vu, il est important de bien choisir les caractéristiques qui vont être comparées pour obtenir une méthode aussi robuste que possible.

Lorsque nous procédons à l'enregistrement d'un nouveau sujet, nous obtenons les coordonnées 3D des 25 articulations. Si nous observons la courbe d'une coordonnée au cours du temps et la comparons avec le modèle que nous avons stocké qui correspond à une autre personne (et probablement d'une hauteur différente), la courbe obtenue ne sera pas la même. Par conséquent nous ne pourrons jamais identifier que le mouvement est le même.

Une façon de rendre l'algorithme invariant à l'échelle est d'utiliser les angles au lieu des coordonnées.

Suivant l'idée de Rihawi *et al.* dans [10], nous proposons aussi l'extraction de deux angles θ_{YZ} et θ_{ZX} pour toutes les paires d'articulations.

Pour chaque paire d'articulations, nous considérerons que l'articulation supérieure est celle qui est la plus haute dans le corps (debout et avec les bras vers le bas) et l'inférieure la plus bas. Par exemple, entre l'épaule et le coude, l'épaule sera la supérieure (P_s) et le coude sera l'inférieure (P_i).

$$\begin{aligned} P_s &= (x_s,\, y_s,\, z_s) \\ P_i &= (x_i,\, y_i,\, z_i) \end{aligned} \tag{4.11}$$

Le calcul de ces deux angles est donné par les équations suivantes :

$$\theta_{YZ} = arctan\left(\frac{y_i - y_s}{\sqrt{(x_i - x_s)^2 + (z_i - z_s)^2}}\right) \tag{4.12}$$

$$\theta_{ZX} = arctan\left(\frac{x_i - x_s}{z_i - z_s}\right) \tag{4.13}$$

Nous pouvons voir un exemple de ces deux angles dans la figure 4.2 où le point inférieur (P_i) correspond au coude gauche et le point supérieur (P_s) à l'épaule gauche.

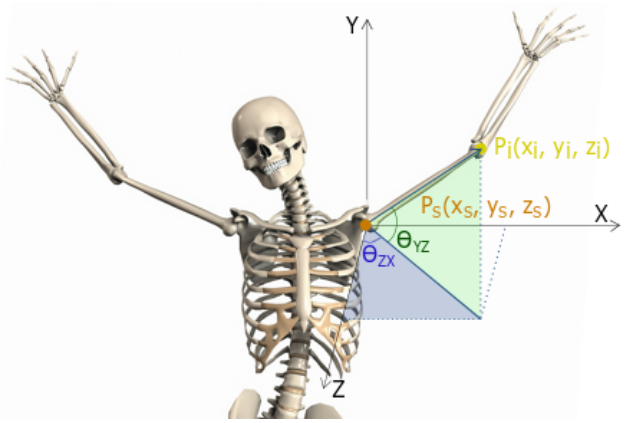


FIGURE 4.2 – Angles θ_{YZ} et θ_{ZX} entre le coude gauche (P_i) et l'épaule gauche (P_s).

Avec ces deux angles, nous avons réussi à extraire des caractéristiques invariantes à l'échelle de la personne, mais pas invariants à sa rotation par rapport à la Kinect, c'est-à-dire des axes étant orientés vers la caméra. En effet, si nous sommes dans la position de l'image à gauche de la figure 4.3, nous voulons avoir $\theta_{ZX} = 0°$ (considérant P_i comme le coude droit et P_s comme l'épaule droite). En revanche, si nous sommes dans la position de l'image à droite nous aurons également un angle $\theta_{ZX} = 0°$.

Afin de rendre les caractéristiques invariantes à la rotation, nous calculerons l'angle de notre torse avec l'axe Z, c'est-à-dire l'angle de notre corps par rapport à la direction de la caméra. Nous calculons cet angle de la manière suivante :

$$\theta_{Torse} = arctan\left(\frac{z_{épaule-droite} - z_{épaule-gauche}}{x_{épaule-gauche} - x_{épaule-droite}}\right) \tag{4.14}$$

$\theta_{Torse} = 0\ °$ $\theta_{Torse} = 31.47°$

FIGURE 4.3 – Différentes positions du torse par rapport à la Kinect.

Dans le cas de l'image à droite de la figure 4.3 l'angle du torse par rapport à la Kinect est de 31.47°, nous devrons ajouter cet angle θ_{Torse} à l'angle θ_{ZX} ce qui est équivalent à faire pivoter le système de coordonnées autour de l'axe Y.

4.2.2 Première approche

Une fois que nous avons choisi les caractéristiques que nous allons comparer, l'étape suivante est de choisir l'approche que nous allons utiliser pour comparer les courbes.

D'une part avec l'algorithme DTW, qui nous permet de calculer la distance minimale entre deux courbes, nous devons d'abord connaître la longueur des deux courbes que nous voulons comparer, mais ce n'est pas notre cas. Nous avons un motif (dont nous connaissons la longueur) et nous voulons savoir si dans le signal que nous avons enregistré (1000 images), il y a une partie qui se rapproche de notre modèle.

D'autre part l'algorithme de correspondance de courbe expliqué dans la section 4.1.2, permet de trouver la portion de courbe qui ressemble le plus au motif de référence. Pour cette raison, nous avons pensé que cet algorithme résoudrait mieux notre problème.

Cependant, cet algorithme n'est pas adapté à notre problème. Il faut remarquer qu'il est pensé pour faire correspondre deux courbes sous une transformation de similarité, c'est-à-dire sous une translation, rotation ou bien sous un changement d'échelle uniforme. Au contraire, dans notre cas, les courbes sont plus sujettes à des transformations dues à la différence d'exécution de chaque sujet, mais aussi au bruit introduit par le manque de précision du capteur, et donc nous n'obtenons pas les résultats souhaités.

4.2.3 Deuxième approche

Dans le cas de cette deuxième approche, nous avons décidé d'utiliser l'algorithme DTW, car selon la littérature, ce sont les meilleurs résultats obtenus si notre objectif est la comparaison des courbes.

Cependant, comme expliqué dans la section précédente, le principal problème que nous avons est de trouver la longueur du motif que nous recherchons. Afin, de ne pas utiliser une recherche exhaustive, nous avons décidé d'utiliser l'autocorrélation.

L'autocorrélation est la corrélation croisée du signal avec lui-même, c'est-à-dire qu'elle mesure la similitude avec lui-même et nous permet donc de trouver des motifs répétitifs, ainsi que sa période.

Nous voulons donc connaître la longueur de mouvements répétitifs : cela revient à connaître la période du signal. Nous devons garder à l'esprit que la même personne peut effectuer le même mouvement à des vitesses différentes, c'est pourquoi nous devons trouver les différentes longueurs du mouvement.

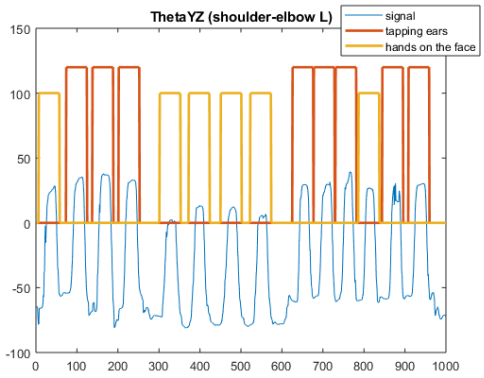
Pour ce faire, nous avons parcouru le signal avec une fenêtre de 200 images et pour chaque nouvelle itération, nous déplaçons la fenêtre 50 images. Pour chaque fragment, nous allons calculer l'autocorrélation pour trouver s'il y a un motif qui est répété, s'il y a un motif que se répète nous enregistrons la période. Une fois le signal est parcouru, nous aurons un vecteur des différentes périodes.

Ensuite, nous utiliserons l'algorithme DTW pour calculer la distance à notre motif. Mais, au lieu de parcourir le signal avec une fenêtre de longueur fixe, nous allons tester les différentes longueurs obtenues.

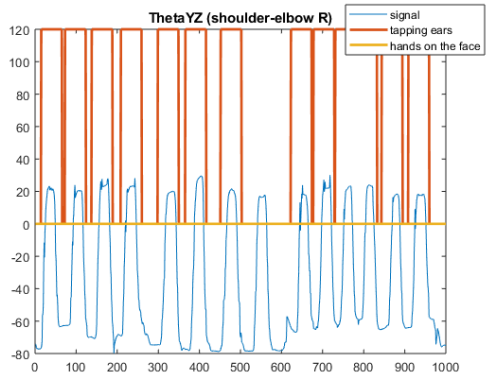
Ci-dessous nous détaillons les différentes étapes suivies ainsi que les différents paramètres utilisés pour la détection du mouvement des mains sur le visage et ainsi que le mouvement des mains sur les oreilles, voir aussi le résumé dans la figure 4.7.

1. **Première étape :** grâce à l'autocorrélation nous calculerons la période des mouvements répétitifs que nous trouvons.
2. **Deuxième étape :** pour ces deux mouvements les angles que nous avons considérés importants pour la détection, sont les suivants :
   - L'angle du vecteur qui joint l'épaule au coude par rapport au plan YZ du bras gauche.
   - L'angle du vecteur qui joint l'épaule au coude par rapport au plan YZ du bras droit.
   - L'angle du vecteur qui joint le coude à la main par rapport au plan YZ du bras gauche.
   - L'angle du vecteur qui joint le coude à la main par rapport au plan YZ du bras droit.

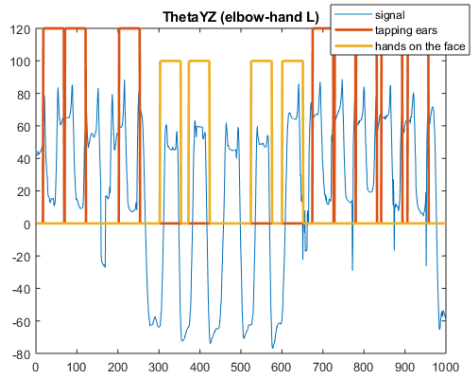
Pour chaque signal de ces quatre angles utilisés, nous appliquerons l'algorithme DTW. La distance DTW sera calculée par rapport au motif des mains sur le visage et le motif des mains couvrant les oreilles et nous prendrons la plus petite distance et en même temps inférieur à 600, c'est-à-dire le mouvement qui ressemble le plus.



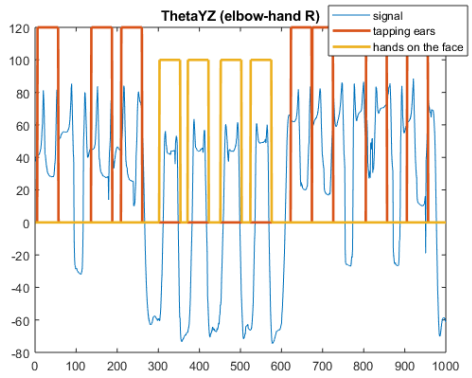
ThetaYZ (épaule-coude G)



ThetaYZ (épaule-coude D)



ThetaYZ (coude-main G)



ThetaYZ (coude-main D)

FIGURE 4.4 – Résultats pour les 4 angles utilisés pour la détection de ce mouvement (G=gauche, D=droit). En orange et avec une valeur de 120, nous représentons les fragments où la distance obtenue avec l'algorithme DTW était inférieure à 600 et plus petite si nous le comparons avec le motif de couvrir les oreilles qu'avec le motif de couvrir le visage avec les mains. En jaune et avec une valeur de 100, les fragments où la distance obtenue était inférieure à 600 et plus petite par rapport au motif de couvrir le visage, en d'autres termes le signal ressemble à ce motif.

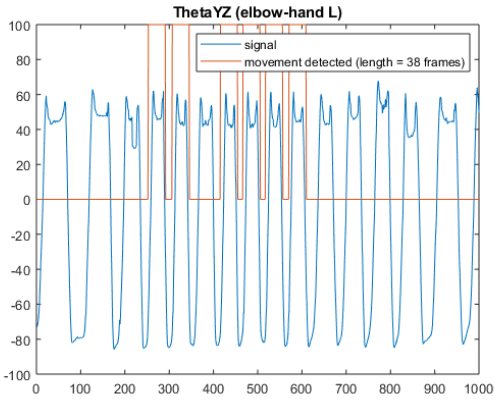
Dans la figure 4.4 nous pouvons voir le signal des quatre angles que nous utilisons pour détecter les deux mouvements recherchés. En bleu nous représentons le signal des différents angles pour la même séquence d'images, de l'image 0 à l'image 300 et de l'image 660 à 1000, le mouvement des mains sur les oreilles a été fait et de l'image 300 à 660 le mouvement des mains sur le visage. En orange et avec une valeur de 120 nous indiquons les images où la distance DTW par rapport au motif du mouvement des mains sur les oreilles est inférieure à 600 et inférieure à la distance DTW par rapport au motif du mouvement des mains sur le visage, c'est-à-dire, les images où nous avons détecté que le mouvement des mains sur les oreilles a été fait. Et en jaune et avec une valeur de 100, nous représentons les images où le mouvement des mains sur le visage a été fait. Si nous regardons la première ligne de la figure 4.4, nous pouvons voir le signal qui correspond à l'angle du vecteur qui joint l'épaule au coude par rapport au plan YZ du bras gauche et droit et nous pouvons vérifier que les signaux sont assez similaires pour les deux mouvements et c'est pourquoi le résultat est parfois faux. En revanche, dans la deuxième ligne, nous pouvons voir le signal qui correspond à l'angle du vecteur qui joint le coude à la main du bras gauche et droit par rapport au plan YZ et nous voyons que la détection est plus proche de la solution attendue. En combinant les résultats des quatre angles (par sommation), nous déciderons si le mouvement a été les mains sur le visage ou les mains sur les oreilles. Grâce à cette utilisation des différents angles, nous pouvons arriver à la conclusion que le mouvement fait dans le fragment jusqu'à l'image 300 et de l'image 660 à 1000 correspond au mouvement des mains sur les oreilles et de l'image 300 à 660 correspond au mouvement des mains sur le visage.

3. **Troisième étape :** jusqu'à présent, nous avons le résultat pour une longueur de la période, dans cet exemple que nous venons de montrer la longueur est de 50 images. L'étape suivante consiste à répéter la deuxième étape pour chacune des périodes trouvées.

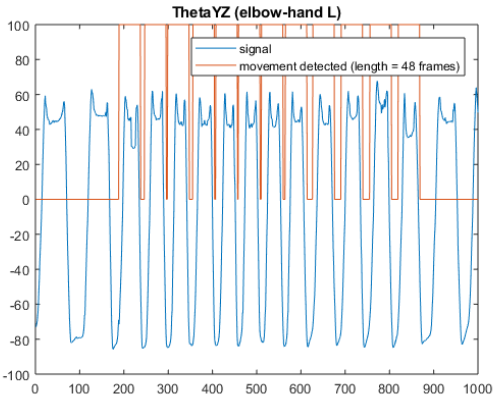
   Dans la figure 4.5 un autre exemple est donné, le signal bleu correspondant à l'angle du vecteur qui joint le coude à la main par rapport au plan YZ du bras gauche. Nous avons représenté en orange et avec une valeur de 100, les fragments où la distance obtenue par l'algorithme DTW par rapport au motif du mouvement couvrant le visage avec les mains est inférieur à 600 et aussi inférieur à la distance par rapport au motif du mouvement couvrant les oreilles. C'est-à-dire, dans ces fragments nous dirons que le mouvement couvrant le visage avec les

mains a été fait.

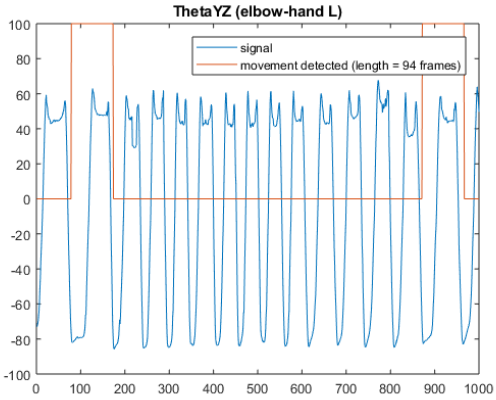
Nous voyons 3 des longueurs testées : 38, 48 et 94. Comme nous pouvons le voir, cela nous a permis de détecter le mouvement même si celui-ci a été effectué à une vitesse différente.



Durée de la période = 38



Durée de la période = 48



Durée de la période = 94

FIGURE 4.5 – Résultats de la détection du mouvement de couvrir le visage avec les mains pour trois longueurs de période différentes. Notez que dans cet exemple est affiché seulement un angle, concrètement l'angle du vecteur qui joint le coude à la main par rapport au plan YZ.

4. **Quatrième étape :** finalement, en utilisant les résultats de la détection des différentes périodes, nous décidons lequel des deux mouvements a été effectué. Dans la figure 4.6 nous pouvons voir un exemple de résultat de la détection de ces deux mouvements après avoir utilisé les résultats des différentes périodes.

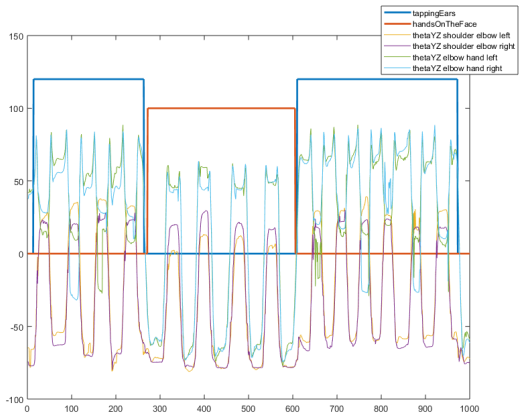


FIGURE 4.6 – Résultats de la détection des deux mouvements en utilisant les résultats des quatre angles ainsi que les différentes longueurs. En bleu et avec une valeur de 120 les fragments où le mouvement de couvrir les oreilles ont été détectés. En orange et avec une valeur de 100 les fragments où le mouvement de couvrir le visage ont été détectés.

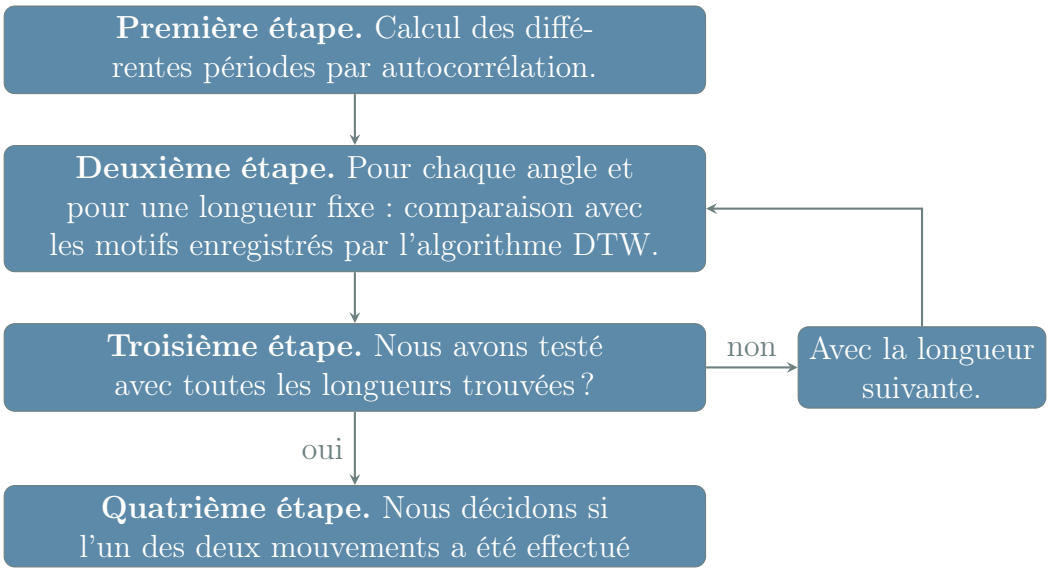


FIGURE 4.7 – Résumé des différentes étapes de cette méthode, pour la détection du mouvement stéréotypé de couvrir les oreilles avec les mains et pour le mouvement de couvrir le visage avec les mains.

4.2.3.1 Limites

Dans cette approche nous avons réussi à détecter les deux mouvements stéréotypes, même s'ils ont été effectués à des vitesses différentes. Cependant, nous supposons que le sujet effectue chaque mouvement à plusieurs reprises.

Dans le cas supposé, où le sujet fait une seule fois ou dans le cas où la pause entre les mouvements est importante, l'algorithme ne détectera pas la période et par conséquent ne détectera pas le mouvement.

4.2.4 Approche finale

Comme nous l'avons vu dans la section précédente, grâce à l'algorithme DTW, nous obtenons des résultats assez satisfaisants, mais d'une part nous devons tester l'algorithme DTW pour chacune des longueurs trouvées. Et d'un autre côté, comme nous l'avons également expliqué, nous ne pourrons détecter les mouvements que s'ils sont faits à plusieurs reprises.

Afin d'améliorer ces aspects, dans cette section nous allons donc introduire une nouvelle méthode qui nous permettra également de détecter des mouvements isolés en utilisant une modification du DTW classique pour trouver le motif recherché dans un signal plus long, cette modification est proposée par Müller dans [9].

Si nous revenons aux deux séries dont nous avons parlé dans la section 4.1.3 :

$$\begin{aligned} X &:= (x_1,\, x_2,\, ...\, x_i,\, ...\, x_N) \\ Y &:= (y_1,\, y_2,\, ...\, y_j,\, ...\, y_M) \end{aligned} \tag{4.15}$$

Et que nous supposons que M est plus grand que N, ce qui est notre cas, voir un exemple sur la figure 4.8. Où nous pouvons voir en bleu, le signal d'un nouveau sujet, correspondant à l'angle du vecteur qui joint le coude à la main par rapport au plan YZ du bras droit. En rose nous avons le signal du même angle correspondant au modèle enregistré du mouvement couvrant le visage avec les mains.

Ce que nous cherchons donc c'est trouver une sous-séquence :

$$Y(a^* : b^*) = (y_{a^*}, y_{a^*+1}, ..., y_{b^*}) \text{ avec } 1 \leq a^* \leq b^* \leq M \tag{4.16}$$

Le but est alors de trouver a^* et b^* pour lesquels l'alignement entre X et Y est optimal. Comme nous l'avons expliqué précédemment dans la section 4.1.3, la condition aux limites force le début de l'alignement au premier élément des deux séries et se termine avec le dernier élément des deux séries. Au lieu de cela, ce que Müller propose ne tient pas compte de cette condition, les étapes à suivre sont les suivantes :

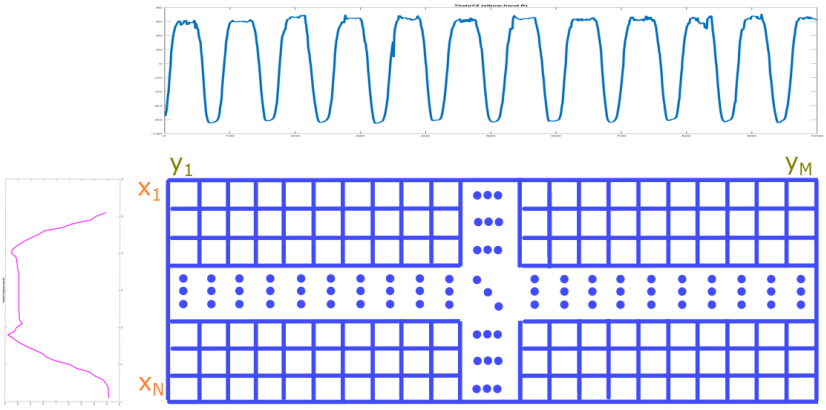


FIGURE 4.8 – Matrice entre les deux signaux : en bleu le signal correspondant à l'angle du vecteur qui joint le coude à la main par rapport au plan YZ du bras droit et en rose le signal (du même angle) du motif enregistré correspondant au mouvement de couvrir le visage avec les mains.

1. Tout d'abord nous calculons la **matrice distance** entre les deux signaux comme nous l'avons faite dans l'approche DTW classique, voir l'équation 4.2.
2. L'étape suivante est le calcul de la **matrice de coûts** cumulés (nous l'avons appelée c dans la section 4.1.3). Müller propose ici un changement dans le calcul de cette matrice. La première ligne de la matrice c avec une longueur de M (voir figure 4.9), ne tiendra pas compte de la valeur précédente de la matrice des coûts cumulés, c'est-à-dire qu'elle sera simplement égale à la distance :

$$c(i,j) = d(1,j) \quad \text{si } i = 1 \text{ et } j \neq 1 \tag{4.17}$$

   Le reste de la matrice nous le calculons comme nous l'avons fait dans la méthode classique, voir l'équation 4.3.
   Notez que la représentation de la matrice des coûts cumulés dans la figure 4.9 a été représentée en échelle de gris, le blanc correspond à un coût élevé et le noir à un faible coût.
3. L'étape suivante consistera à calculer b^* défini comme :

$$b^* = \underset{b \in [1:M]}{\arg\min}\, c(N,b) \tag{4.18}$$

   Autrement dit, à partir de la matrice de coûts cumulés de la figure 4.9, la valeur correspondant à b^* sera le minimum de la dernière ligne de cette matrice.

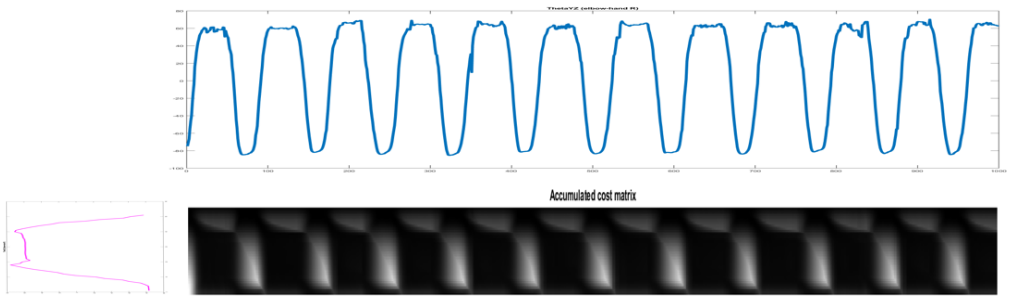


FIGURE 4.9 – Matrice de coûts cumulés modifiée. Le blanc correspond à un coût élevé et le noir à un faible coût.

4. Finalement, nous traçons le **chemin de déformation** optimal à partir de $p_L = (N, b^*)$ sans forcer a^* à être 1, c'est-à-dire, jusqu'à $p_1 = (1, a^*)$. Nous allons choisir a^* en suivant les étapes suivantes : une fois que nous atteignons la première ligne, si l'élément précédent, c'est-à-dire $c(1, a-1)$, est plus grand que $c(1, a)$ nous prendrons a comme a^*. Nous pouvons voir dans la figure 4.10 le chemin optimal de $p_L = (N, b^*)$ à $p_1 = (1, a^*)$, affiché en blanc sur la matrice des coûts.

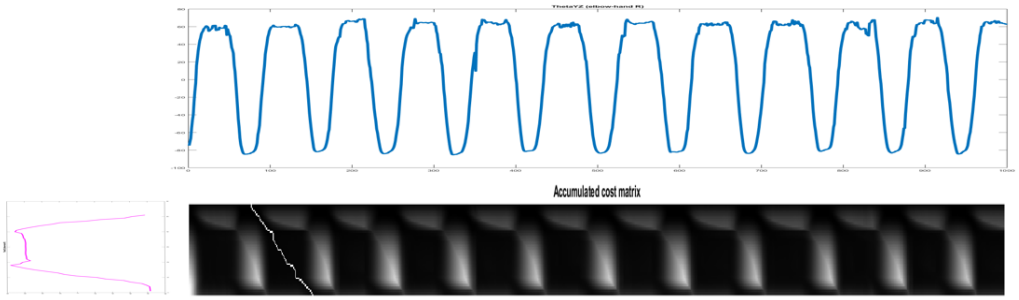


FIGURE 4.10 – Chemin optimal (chemin dessiné en blanc), c'est-à-dire le meilleur alignement entre le motif (signal en rose) et un fragment du signal du sujet (signal bleu).

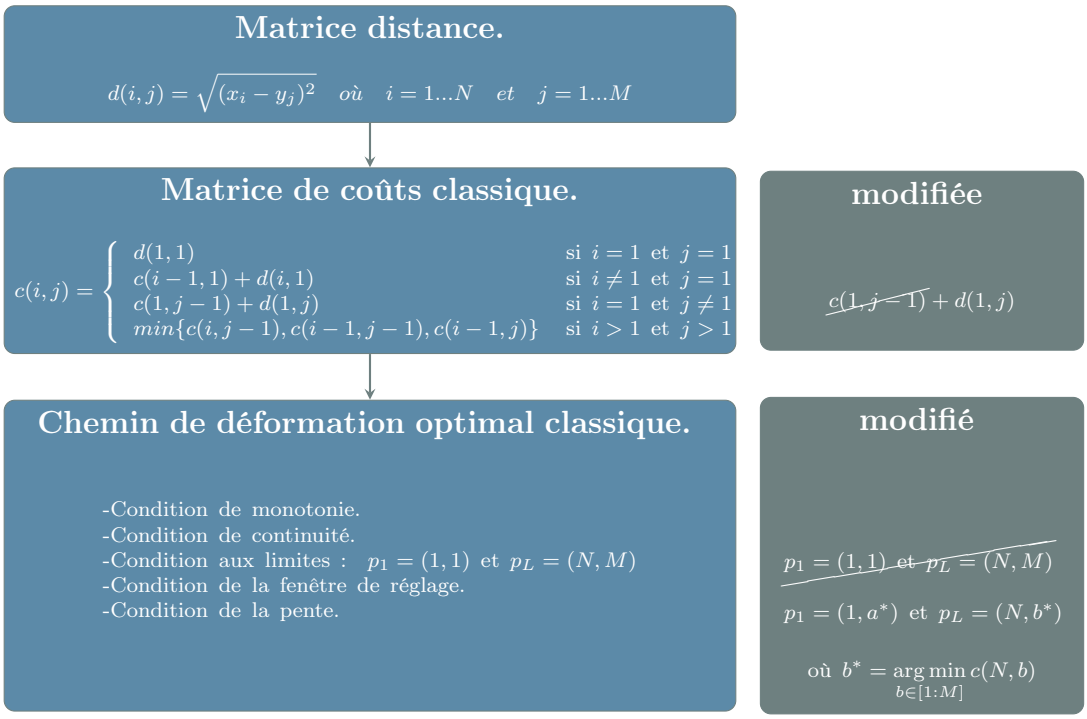


FIGURE 4.11 – En bleu les différentes étapes de l'algorithme DTW classique. En gris les modifications proposées par Müller dans [9].

Dans la figure 4.11, nous avons résumé, en bleu, les différentes étapes de l'algorithme DTW classique et en gris les modifications proposées par Müller dans [9], qui viennent d'être expliquées.

Pour aller plus loin, Müller propose aussi une approche pour détecter toutes les sous-séquences qui ressemblent au motif recherché. La procédure est la suivante, au lieu de ne prendre que le minimum de la dernière ligne de la matrice des coûts cumulés (voir figure 4.9), tous les minima inférieurs à un seuil seront considérés comme des b^*. Nous pouvons voir dans la figure 4.12 les différents minima correspondant aux b^*.

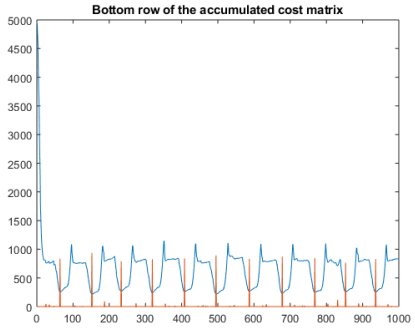


FIGURE 4.12 – Dernière ligne de la matrice des coûts cumulés. Les différents minimums correspondent aux b^* recherchées.

Notez que pour ne pas prendre plusieurs sous-séquences qui ne diffèrent que de quelques images, nous choisirons les b^* comme minimums locaux, c'est-à-dire que le prochain b^* devrait être au-delà du maximum local à droite du b^* courant et le précédent b^* devrait être avant le maximum local à gauche du b^* courant.

Pour trouver les différents a^* nous calculerons les chemins optimaux correspondant à partir de chaque b^*, comme nous venons d'expliquer. Nous pouvons voir dans la figure 4.13 les différents chemins optimaux représentés, correspondant aux différents motifs trouvés.

4.2.4.1 Mouvements à détecter

Dans cette sous-section, nous expliquerons plus en détail les angles utilisés pour la détection de chacun des quatre mouvements stéréotypés, ainsi que les différents paramètres utilisés.

1. Mains sur le visage : pour la détection de ce mouvement en utilisant cette méthode, nous avons décidé d'utiliser l'angle du vecteur qui joint le coude à la main par rapport au plan YZ du bras gauche et du bras

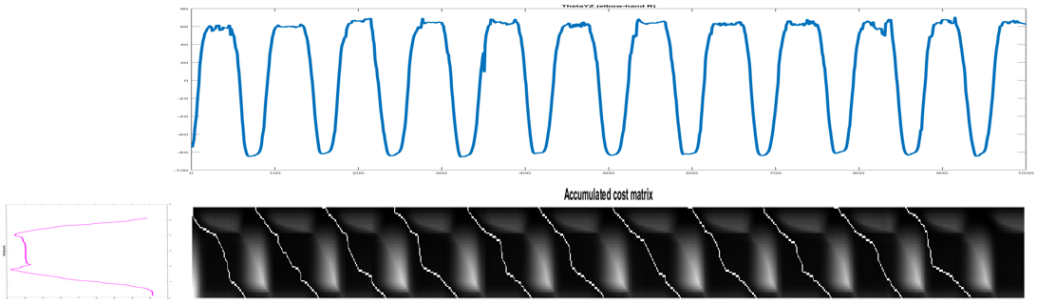


FIGURE 4.13 – Chemins optimaux, en commençant chacun pour un b^*. Chaque chemin optimal montre la correspondance possible entre le motif (signal en rose) et un fragment du signal du sujet (signal en bleu).

droit. Le seuil choisi pour sélectionner les minimums correspondant à ce mouvement est 600.

2. Mains sur les oreilles : pour la détection de ce mouvement, les angles sélectionnés seront l'angle du vecteur qui joint l'épaule au coude par rapport au plan YZ du bras gauche et du bras droit. Le seuil choisi pour sélectionner les minimums correspondant à ce mouvement est de 600.

3. Les coups de tête (*head banging*) : ce mouvement est difficile à détecter car si, par exemple, nous ne prenons que l'angle qui joint la tête au cou et le cou au point entre les épaules, nous prenons le risque de détecter comme un mouvement stéréotypé tout mouvement de la tête en avant même sans l'être. D'autre part, ce mouvement est un mouvement très caractéristique des enfants autistes, c'est pourquoi nous avons jugé important de le détecter. Donc, au lieu d'utiliser un ou deux angles, nous avons décidé d'utiliser 4 angles pour sa détection :
   — L'angle du vecteur qui joint la tête au cou par rapport au plan YZ.
   — L'angle du vecteur qui joint le cou au point entre les épaules par rapport au plan YZ.
   — L'angle du vecteur qui joint le point entre les épaules au point qui se trouve sur la poitrine par rapport au plan YZ.
   — L'angle du vecteur qui joint le point qui se trouve sur la poitrine au point qui est au centre de la hanche par rapport au plan YZ.

   Le seuil choisi pour sélectionner les minimums correspondant à ce mouvement est de 400. Et pour éviter de détecter des faux positifs, c'est-à-dire pour éviter que tout mouvement en avant de la tête soit reconnu comme étant celui d'un autiste, nous déciderons que ce mouvement a été fait, s'il a été détecté dans les 4 angles. Cela signifie

que si le mouvement n'est pas assez marqué, il ne le détectera pas, de cette façon nous ne détecterons pas tout mouvement de la tête comme le mouvement d'un autiste, au prix de perdre des coups de tête moins marquées qui éventuellement sont effectués par un sujet atteint d'autisme.

4. Battement de mains (*hand flapping*) : ce nouveau mouvement est également typique parmi les enfants autistes. Pour détecter ce mouvement, nous utilisons les angles suivants :
   — L'angle du vecteur qui joint le coude à la main du bras droit par rapport au plan YZ.
   — L'angle du vecteur qui joint le coude à la main du bras gauche par rapport au plan YZ.
   — L'angle du vecteur qui joint le coude à la main du bras droit par rapport au plan ZX.
   — L'angle du vecteur qui joint le coude à la main du bras gauche par rapport au plan ZX.

   Le seuil choisi pour ce mouvement est de 200. Comme dans le mouvement de la tête, dans le battement des mains, nous avons également essayé de ne pas avoir de faux positifs, c'est-à-dire que nous préférons ne pas détecter un battement des mains qui devrait être détecté au lieu de détecter quel que soit le mouvement comme un battement de mains, pour cela nous considérons que le battement des mains a eu lieu si le mouvement est détecté dans trois des quatre angles précédemment mentionnés.

4.3 Résultats

Dans cette section, nous présentons les résultats obtenus avec la méthode expliquée dans la section 4.2.4.

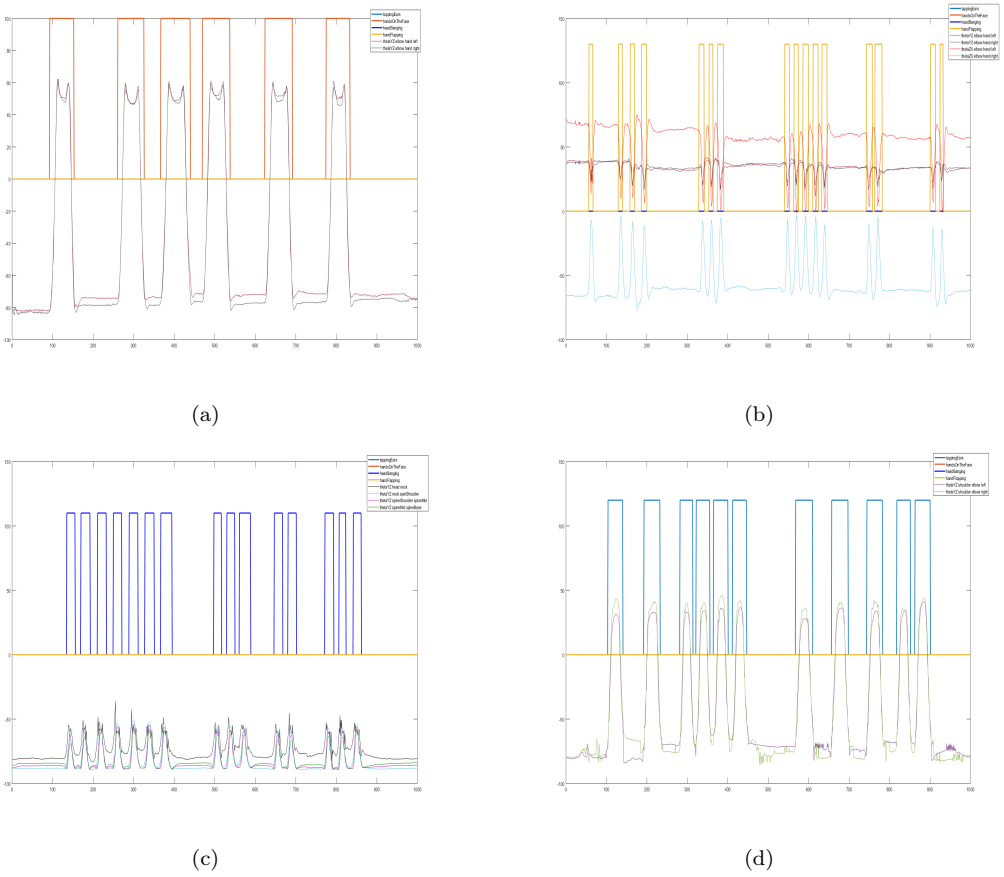
Bien que le but de notre approche soit d'être invariant par rapport au sujet, c'est-à-dire, bien que nous ayons enregistré les modèles avec une personne spécifique, la méthode devrait détecter les mouvements d'une autre personne aussi. De toute façon nous avons commencé par tester la méthode avec la même, c'est-à-dire tout d'abord les séquences avec lesquelles nous testons la méthode ont été faites par la même personne à qui les motifs correspondent.

Dans la figure 4.14, nous pouvons voir les résultats de différentes séquences où l'algorithme a correctement détecté le mouvement effectué. Dans chaque séquence, nous avons réalisé un seul mouvement et pour chaque mouvement nous n'avons représenté que la trajectoire des angles utilisés pour sa détection (voir section 4.2.4.1 pour plus de détails).

Le code couleur utilisé pour indiquer les fragments dans lesquels chaque mouvement a été détecté est le suivant :

— En orange et avec une valeur de 100 : les mains sur le visage.
— En jaune et avec une valeur de 130 : le battement de mains.
— En bleu foncé et avec une valeur de 110 : le balancement du corps.
— En bleu clair et avec une valeur de 120 : les mains sur les oreilles.

Notez que ce code sera utilisé à partir de maintenant dans toutes les figures.



(a) (b)

(c) (d)

FIGURE 4.14 – Résultats de la détection des différents mouvements. Dans chaque figure, nous n'avons représenté que les angles utilisés pour sa détection. (**a**) Mains sur le visage : en orange et avec une valeur de 100, les sections où ce mouvement a été détecté. (**b**) En jaune et avec une valeur de 130 les fragments où le mouvement de battement de mains a eu lieu. (**c**) Les fragments où les coups de tête sont détectés sont représentés en bleu foncé et avec une valeur de 110. (**d**) En bleu clair et avec une valeur de 120, les fragments où le mouvement des mains vers les oreilles a été reconnu.

Nous pouvons voir comment nous avons correctement détecté les différents mouvements. En outre grâce à cette nouvelle méthode, nous sommes même capables de détecter des mouvements isolés, c'est-à-dire que le mouvement sera détecté même s'il n'a pas été fait à plusieurs reprises.

De plus, dans la figure 4.15, nous avons représenté deux séquences où les quatre mouvements ont été effectués et ils ont été correctement détectés. Notez que pour une meilleure visualisation, en ceci nous n'avons pas représenté tous les angles utilisés pour la détection, seulement trois :

— En grenat, l'angle du vecteur qui joint le coude à la main du bras gauche par rapport au plan YZ.
— En noir, l'angle du vecteur qui joint le coude à la main du bras droit par rapport au plan YZ.
— En bleu ciel, l'angle du vecteur qui joint le point entre les épaules au point qui se trouve sur la poitrine par rapport au plan YZ.

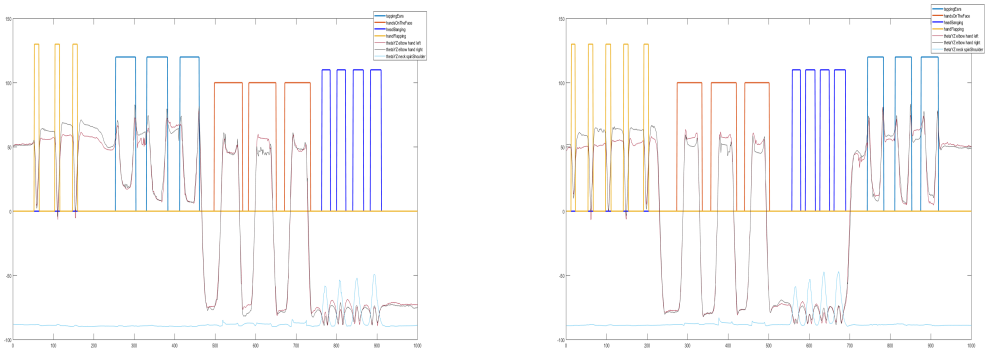


FIGURE 4.15 – Résultats de la détection des quatre mouvements. En orange les mains sur le visage, en jaune le battement des mains, en bleu foncé les coups de tête et en bleu clair les mains sur les oreilles. Notez aussi que nous n'avons représenté que trois des angles utilisés pour la détection.

Comme nous avons vu les résultats avec la même personne sont tout à fait satisfaisants, nous procédons maintenant aux tests avec d'autres sujets.

Dans la figure 4.16, nous pouvons voir la détection correcte des deux mouvements, avec un sujet de 1,75 mètre. Dans la figure de gauche, le mouvement de couvrir le visage a été fait et détecté, les trajectoires des deux angles représentés sont les deux utilisées pour la détection de ce mouvement. De même dans la figure à droite le mouvement de couvrir les oreilles avec les mains a

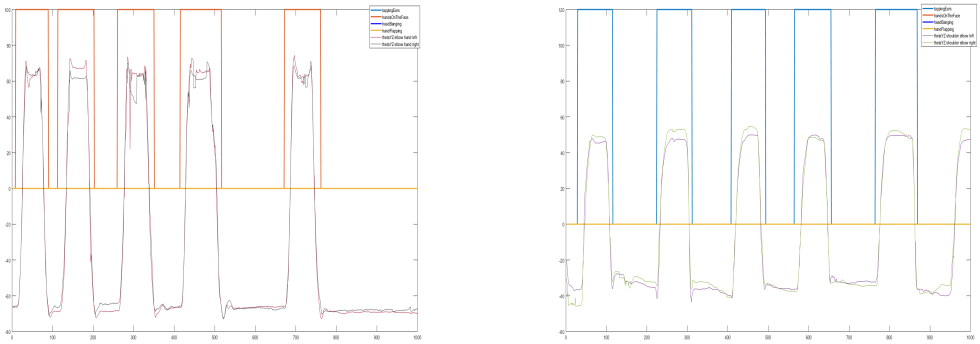


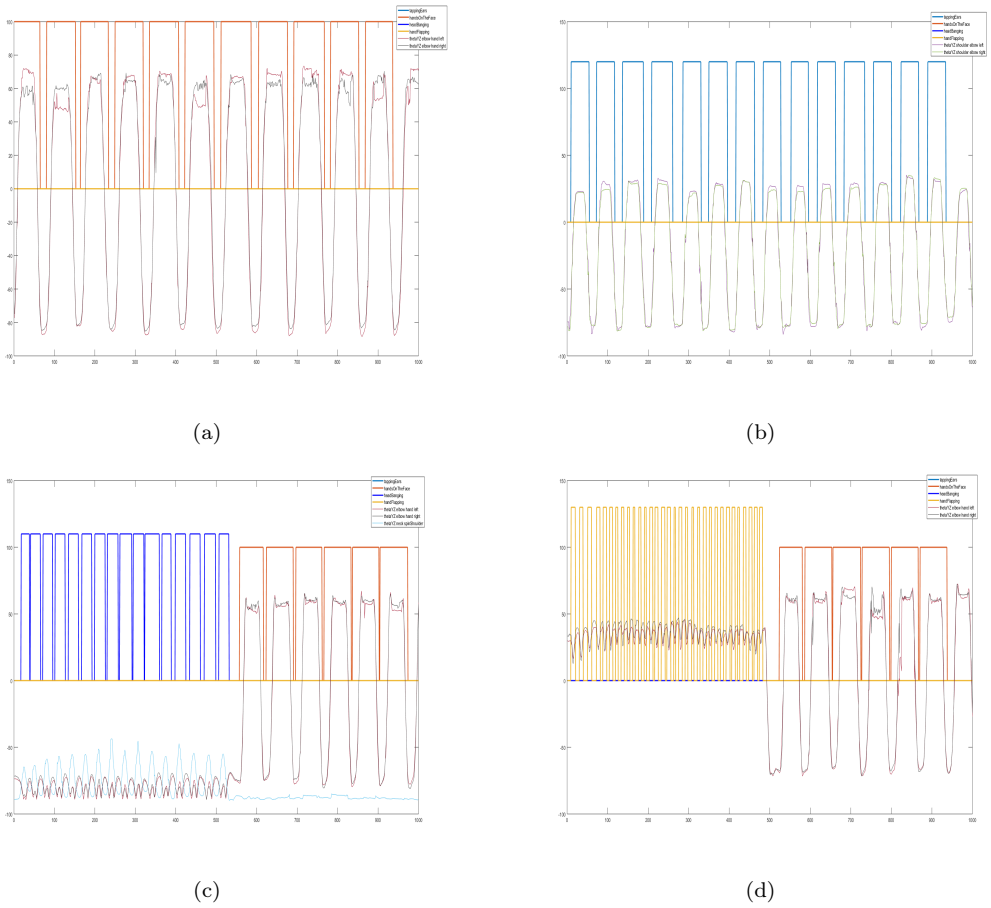
FIGURE 4.16 – Résultats de la détection (sujet de 1.75 m). En en orange le mouvement de couvrir le visage et en bleu clair le mouvement de couvrir les oreilles. Notez que dans chaque figure, seulement les angles utilisés pour la détection de chaque mouvement ont été représentés.

été fait et détecté et également les deux angles utilisés pour la détection sont ceux représentés sur la figure.

Comme le modèle enregistré correspond à une personne qui mesure 1,53 mètre, nous avons trouvé pratique de la tester aussi avec une personne de 1,91 mètre pour vérifier que la méthode est totalement invariante à la taille de la personne. Ainsi, dans la figure 4.17, les résultats des différentes séquences d'une personne de 1,91 mètre, sont affichés. Nous pouvons voir comme nous sommes capables de bien détecter les quatre mouvements.

Notez que dans la figure (**c**) et (**d**) nous n'avons pas représenté tous les angles utilisés pour la détection des mouvements afin d'obtenir une meilleure visualisation.

Pour conclure avec cette méthode, nous pouvons affirmer que nous sommes capables de détecter les quatre mouvements avec assez de précision, même avec des personnes différentes de celle du motif. Or, la difficulté ou la limite de cet algorithme sont les valeurs des seuils, c'est-à-dire le choix du seuil pour décider si la distance est proche ou non du motif.



(a) (b)

(c) (d)

FIGURE 4.17 – Résultats de la détection (sujet de 1.91 m). En bleu foncé les coups de tête, en bleu clair le mouvement de couvrir les oreilles, en orange le mouvement de couvrir le visage et en jaune le battement des mains.

4.3.1 Limites

Les limites de cette approche sont liées à l'utilisation de l'algorithme DTW, c'est-à-dire que cet algorithme fonctionnera lorsque le mouvement effectué soit semblable au modèle que nous avons enregistré. Par exemple, si nous levons les mains vers le haut, nous couvrons nos yeux et nous les relevons à nouveau, le mouvement de couvrir les yeux ne sera pas détecté, puisque notre motif enregistré suppose que nous avons les bras baissés avant de couvrir les yeux.

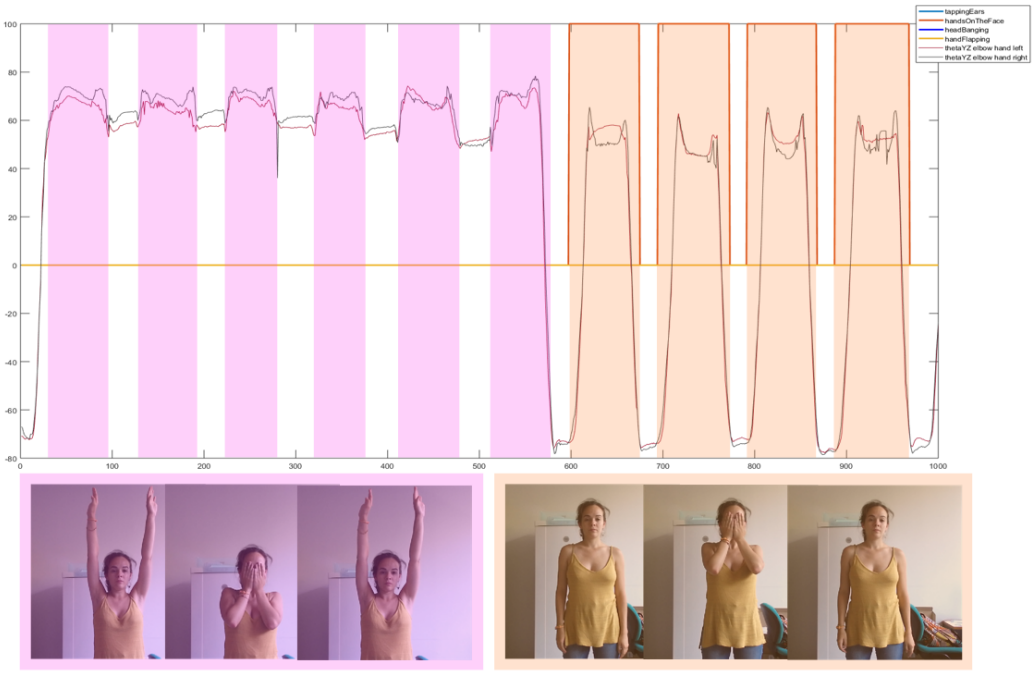


FIGURE 4.18 – En rose les six mouvements de couvrir le visage qui n'ont pas été détectés, faits en suivant le mouvement des images en rose. En orange, les quatre mouvements qui ont été détectés, faits en suivant les images en orange.

Nous avons illustré cet exemple dans la figure 4.18, où les rectangles roses représentent six mouvements de couvrir le visage, en commençant par les bras levés, c'est-à-dire en suivant le mouvement des images en rose au-dessous de la figure. Nous pouvons voir comment il n'a pas été détecté. D'un autre côté, si nous faisons le mouvement comme les images en orange, nous pouvons voir comme elles sont correctement détectées, puisqu'elles correspondent au même mouvement que nous avons enregistré.

5

Conclusion et perspectives

Dans ce travail, une méthode non invasive de détection et reconnaissance de mouvements typiques chez les enfants autistes est proposée. Plus précisément, nous nous sommes concentrés sur la détection de quatre mouvements stéréotypés : le battement des mains, les mains dans le visage, les mains dans les oreilles et les coups de tête ou balancement du corps.

Pour l'extraction des mouvements des personnes, nous avons utilisé un capteur 3D de type Kinect, ce qui nous permet d'obtenir les coordonnées 3D de 25 points du corps humain.

Notre algorithme est invariant à la rotation ainsi qu'à l'échelle de la personne, grâce à l'utilisation des angles au lieu des coordonnées. Et pour l'alignement des différentes trajectoires, nous avons utilisé une modification de l'algorithme DTW classique.

L'efficacité de la méthode proposée a été testé avec la même personne avec qui les motifs ont été enegistrés ainsi qu'avec des personnes de différentes tailles. Mais elle n'a pas pu être testée avec des gestes de personnes autistes.

Nous aimerions pouvoir obtenir les autorisations appropriées pour enregistrer des enfants autistes et tester notre méthode avec ces données. Et donc, être en mesure de vérifier l'objectif de notre approche, c'est-à-dire détecter des troubles du spectre autistique par la vidéo.

La principale limite de l'approche proposée est que l'algorithme fonctionnera lorsque le mouvement effectué soit semblable au modèle que nous avons enregistré. Pour résoudre ce problème, une amélioration possible consisterait à augmenter les motifs enregistrés, c'est-à-dire, au lieu d'avoir un seul modèle pour un mouvement nous pourrions avoir plusieurs types d'exécution du même mouvement.

De même, dans la prochaine étape, nous devrions utiliser plusieurs Kinect afin d'augmenter le champ de vision et éviter les éventuelles articulations cachées, qui posent de problèmes lors de la détection.

Conclusión y perspectivas

En este trabajo, se propone un método no invasivo de detección y reconocimiento de movimientos típicos en niños y niñas con autismo. Específicamente, el proyecto se centra en la detección de cuatro movimientos estereotipados : el aleteo de manos, las manos en la cara, las manos en las orejas y el balanceo del cuerpo.

Para la extracción de los movimientos de las personas, se utiliza un sensor 3D de tipo Kinect, el cual permite obtener las coordenadas 3D de 25 puntos del cuerpo humano.

El método presentado es invariante tanto a la rotación como a la altura de las personas, gracias al uso de los ángulos en lugar de las coordenadas. Y para alinear las diferentes trayectorias se emplea una modificación del algoritmo DTW clásico.

La eficacia del método propuesto ha sido testeada con la misma persona a quien corresponden los patrones grabados, así como con otras personas de diferentes alturas. Este método no ha podido ser testeado con gestos de personas autistas.

El próximo paso de este trabajo consistiría en obtener los permisos necesarios para grabar a niños y niñas autistas y testear este método con sus datos. Y así, poder verificar el objetivo de este proyecto, es decir, detectar trastornos del espectro autista a través del video.

La principal limitación del método propuesto es que el algoritmo funcionará cuando el movimiento sea similar al modelo que hemos guardado precedentemente. Para resolver este problema, una posible mejora sería aumentar los patrones almacenados, es decir, en lugar de tener un patrón único para un movimiento, se podrían tener varios tipos de ejecución del mismo movimiento.

Del mismo modo, en la próxima etapa, se deberían utilizar varios sensores Kinect para aumentar el campo de visión y evitar las posibles articulaciones ocultas, las cuales causan problemas durante la detección.

Bibliographie

[1] American Psychiatric Association. *Diagnostic And Statistical Manual of Mental Disorders : DSM-5*. Washington, D.C. : American Psychiatric Association, c2013., 2013.

[2] Zhang Chenyang and Tian Yingli. RGB-D camera-based daily living activity recognition. *J Comput Vis Image Process.*, 2(4) :12, 2012.

[3] Mingzhu Cui, John C. Femiani, Jiuxiang Hu, Jiuxiang Hu, and Anshuman Razdan. Curve matching for open 2D curves. *Pattern Recognition Letters*, 30(1) :1–10, Jan 2009.

[4] Nuno Gonçalves, José L. Rodrigues, Sandra Costa, and Filomena Soares. Automatic detection of stereotyped hand flapping movements : Two different approaches. In *2012 IEEE RO-MAN : The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 392–397, Sept 2012.

[5] Ulf Großekathöfer, Nikolay V. Manyakov, Vojkan Mihajlovic, Gahan Pandina, Andrew Skalkin, Seth Ness, Abigail Bangerter, and Matthew S. Goodwin. Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. *Frontiers in Neuroinformatics*, 2017.

[6] Jordan Hashemi, Thiago Vallin Spina, Mariano Tepper, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, and Guillermo Sapiro. Computer vision tools for the non-invasive assessment of autism-related behavioral markers. *CoRR*, abs/1210.7014, 2012.

[7] Maha Jazouli, Aicha Majda, Djamal Merad, Rachid Aalouane, and Arsalane Zarghili. Automatic detection of stereotyped movements in autistic children using the kinect sensor. *Journal of Biomedical Engineering and Technology*, April 2017.

[8] Sohaib Laraba, Med Brahimi, Joëlle Tilmanne, and Thierry Dutoit. 3D skeleton-based action recognition by representing motion capture sequences as 2D-RGB images. *Computer Animation and Virtual Worlds*, 28, May 2017.

[9] Meinard Müller. Dynamic time warping. In *Information Retrieval for Music and Motion*, pages 69–84, 2007.

[10] Omar Rihawi, Djamel Merad, and Jean-Luc Damoiseaux. 3D-AD : 3D-Autism Dataset for repetitive behaviours with kinect sensor. In *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, 2017.

[11] Santiago Riofrío, David Pozo, Jorge Rosero, and Juan Vásquez. Gesture recognition using dynamic time warping and kinect : A practical approach. In *2017 International Conference on Information Systems and Computer Science (INCISCOS)*, pages 302–308, Nov 2017.

[12] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. In *IEEE International Conference on Transactions on Acoustics, Speech, and Signal Processing*, volume 26, pages 43 – 49, Mar 1978.

[13] Samsu Sempena, Nur Ulfa Maulidevi, and Peb Ruswono Aryan. Human action recognition using dynamic time warping. In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, pages 1–5, Jul 2011.

[14] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. Gestures as point clouds : A $P Recognizer for user interface prototypes. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, pages 273–280. ACM, 2012.