REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMERI DE TIZI-OUZOU

FACULTÉ DE GÉNIE ÉLECTRIQUE ET D'INFORMATIQUE

DÉPARTEMENT ÉLÉCTRONIQUE

Rapport

Système de Reconnaissance de Gestes pour la Commande de Dispositifs

MODULE : Recherche documentaire et conception de mémoires

SECTION : M2 Systèmes embarquée

EMAIL : alem.21012000@gmail.com

ANNÉE UNIVERSITAIRE: 2023 – 2024

NOMS:

- ABDEMEZIEM FARID
- ALEM ABDERRAHMANE
- KLOUL MASSINISSA

Sommaire

Table de figures

Introduction

Introduction

Il ne fait désormais plus aucun doute que l'informatique représente la révolution la plus importante et la plus innovante qui a marqué la vie de l'humanité en ce siècle passé. En effet, loin d'être un phénomène de mode éphémère, ou une tendance passagère, l'informatique vient apporter de multiples conforts à notre mode de vie.

Aucun domaine n'est resté étranger à cette stratégie qui offre tant de services aussi bien pour l'entreprise ou l'administration que pour de simples utilisateurs.

Parmi les technologies informatiques développées ces dernières années, la vision par La vision artificielle, également connue sous le nom de vision par ordinateur, est une discipline de l'intelligence artificielle (IA) qui vise à permettre aux machines de percevoir et d'interpréter visuellement le monde qui les entoure. Inspirée par le fonctionnement du système visuel humain, la vision artificielle cherche à doter les ordinateurs de la capacité de comprendre, analyser et interpréter des informations visuelles.

Cette branche de l'informatique repose sur des algorithmes complexes et des modèles mathématiques pour traiter des données visuelles telles que des images ou des vidéos. Elle englobe un large éventail d'applications, allant de la reconnaissance d'objets et de visages à la détection de mouvements, la reconnaissance des gestes, en passant par la segmentation d'images et la compréhension de scènes complexes.

La reconnaissance des gestes permet aux machines de percevoir et d'interpréter visuellement le monde qui les entoure, en particulier les actions humaines. Elle vise à doter les systèmes informatiques de la capacité de comprendre et d'interpréter des informations visuelles de manière similaire à la vision humaine.

Notre travail consiste à développer une application en langage python qui interprète les techniques de reconnaissance à fin de commander des dispositifs ou des appareils.

L’objectif doit permettre de :

- Acquisition de séquences vidéo à partir d'une caméra de smartphone ou une webcam.
- Conversion en niveaux de gris.
- Détection et reconnaissance de gestes de la main pour contrôler la souris du PC

Chapitre 1: Vision Artificielle

Chapitre 1: Vision Artificielle

.I Préambule

La vision artificielle, est une discipline de l'intelligence artificielle qui vise à permettre aux machines de percevoir, interpréter et comprendre visuellement le monde qui les entoure. Elle s'inspire de la capacité humaine à comprendre et interpréter les informations visuelles provenant de l'environnement.

.II Définitions

.II.1 Vision Artificielle

La vision artificielle se réfère à la capacité des machines à interpréter et comprendre des informations visuelles. Cela inclut la reconnaissance d'objets, la détection de formes, la segmentation d'images, etc. [1]

.II.2 Traitement d'Images

Il s'agit de la manipulation d'images numériques pour améliorer leur qualité, extraire des informations ou effectuer des opérations spécifiques tell que le filtrage. [2]

.III Principes Fondamentaux

.III.1 Extraction de Caractéristiques

La vision artificielle implique souvent l'extraction de caractéristiques à partir d'images. Cela consiste à identifier des éléments clés ou des motifs significatifs dans les données visuelles.

.IV Classification

Attribuer une étiquette ou une catégorie à une image en fonction de son contenu. Les réseaux de neurones sont couramment utilisés pour cette tâche. [1]

.IV.1 Détection d'Objets

Identifier et localiser les objets spécifiques dans une image. Les algorithmes de détection d'objets peuvent détecter la présence et la position d'objets multiples.

.V Techniques utilisés

.V.1 Apprentissage Profond (Deep Learning)

.V.1.a Réseaux de Neurones Convolutifs (CNN)

Ces réseaux sont spécialement conçus pour traiter des données spatiales telles que des images. Ils utilisent des filtres convolutifs pour détecter des motifs locaux et hiérarchiques dans les images.

.V.1.b Réseaux de Neurones Récursifs

Utilisés pour la reconnaissance d'objets dans des contextes plus complexes, ces réseaux peuvent capturer des relations à long terme et des dépendances temporelles (des rumeurs). [1]

.V.2 Segmentation d'Images

.V.2.a Segmentation Sémantique

Attribue des étiquettes à chaque pixel d'une image, permettant une compréhension détaillée des différentes régions. Par exemple, la segmentation peut distinguer les différentes parties d'une image médicale. [2]

.V.2.b Segmentation Instance

Identifie et isole individuellement chaque instance d'un objet dans une image, utile pour la détection et la compréhension précise des objets multiples. [2]

.V.3 Traitement du Flux Vidéo

.V.3.a Suivi d'Objets

Suit la trajectoire d'objets spécifiques dans une séquence d'images. Cela est crucial pour des applications telles que la surveillance vidéo et la réalité augmentée. [2]

.V.3.b Reconnaissance d'Activités

Analyse les schémas d'activités dans une séquence vidéo pour identifier des actions spécifiques, comme la marche, la course ou d'autres comportements humains.

.VI Applications Spécifiques

.VI.1 Reconnaissance Faciale

Utilise des algorithmes de vision artificielle pour identifier et vérifier l'identité des individus en se basant sur des caractéristiques faciales uniques. [1]

.VI.2 Conduite Autonome

Intègre des systèmes de vision artificielle pour permettre aux véhicules de détecter et de réagir aux éléments de la route, tels que les autres véhicules, les piétons et les panneaux de signalisation. [1]

.VI.3 Médecine et Imagerie Médicale

La vision artificielle est utilisée pour l'analyse d'images médicales, facilitant la détection précoce des maladies et l'assistance aux chirurgies. [1]

.VI.4 Reconnaissance de Gestes

Permet de détecter et interpréter les mouvements du corps humain pour comprendre les gestes et les actions. Cette application est utilisée dans des domaines tels que la réalité virtuelle, les interfaces homme-machine et la surveillance intelligente. [2]

.VII Discussion

L'avenir de la vision artificielle promet des innovations continues, alimentées par la recherche en intelligence artificielle. En somme, la vision artificielle ouvre la voie à un monde où les machines peuvent non seulement voir, mais aussi comprendre et interagir de manière intelligente avec leur environnement visuel. [3]

Méthodes et état de l’Art

Méthodes et état de l’Art

.I Préambule

L'apprentissage profond, une discipline qui utilise des réseaux neuronaux complexes pour extraire des informations riches à partir de données visuelles. Dans cette exploration, nous plongerons dans l'univers captivant de l'apprentissage profond en mettant en lumière quatre modules emblématiques qui illustrent la puissance de cette approche la Reconnaissance Faciale, le Suivi de Main, et la Détection d'Objet, ainsi que la Segmentation Sémantique. [5]

.II MediaPipe

MediaPipe est une bibliothèque open source développée par Google Research qui permet le développement d'applications de vision par ordinateur en temps réel. Elle offre des modules prêts à l'emploi pour diverses tâches, ce qui simplifie le processus de création d'applications de suivi de mouvements, de reconnaissance faciale. [5]

.III Fonctionnement de MediaPipe

.III.1 Modules

Pour Détecter les parties du corps MediaPipe utilise des modules utilisent des modèles d'apprentissage profond pour effectuer des tâches telles que la détection de pose, la détection de mains, la détection faciale...etc

.III.2 Modelés pré-entrainer

.III.2.a Définition

Les "modules d'apprentissage" se réfèrent à des composants logiciels qui encapsulent des modèles d'apprentissage automatique ou profond pré-entraînés. Ces modèles ont été entraînés sur d'énormes ensembles de données pour effectuer des tâches spécifiques telles que la détection d'objets, la reconnaissance de visages, la segmentation d'images, etc. Ces modules sont souvent utilisés dans des bibliothèques ou des frameworks de vision par ordinateur pour fournir des fonctionnalités prêtes à l'emploi aux développeurs, évitant ainsi la nécessité d'entraîner des modèles à partir de zéro. [5]

.III.2.b Réseaux de Neurones Profonds (Deep Neural Networks)

Les modules d'apprentissage de MediaPipe sont souvent basés sur des réseaux de neurones profonds. Ces réseaux sont des architectures composées de nombreuses couches (d'où le terme "profond") qui apprennent des représentations hiérarchiques des données d'entrée.

Entraînement sur de Grandes Quantités de Données, Les modèles sont entraînés sur des ensembles de données massifs et diversifiés pour garantir qu'ils peuvent généraliser à différentes situations. Par exemple, un modèle de détection de mains peut être entraîné sur des milliers d'images et de vidéos contenant des mains dans diverses poses et conditions d'éclairage. [5]

.III.2.c Tâches Spécifiques

Chaque module d'apprentissage est conçu pour effectuer une tâche spécifique. Par exemple, un module peut être entraîné pour détecter la pose du corps, un autre pour détecter des objets en 2D, et un autre pour la segmentation sémantique. [5]

.III.2.d Prêt à l'Emploi

Les modules d'apprentissage de MediaPipe sont généralement fournis sous forme de modèles pré-entraînés. Cela signifie qu'ils sont prêts à être utilisés dans des applications nécessité un nouvel entraînement sur des données supplémentaires. [5]

.III.2.e Intégration dans des Applications :

Ces modules d'apprentissage sont intégrés dans la bibliothèque MediaPipe, fournissant ainsi une interface conviviale pour les développeurs afin qu'ils puissent facilement tirer parti de ces fonctionnalités dans leurs propres applications. [4]

.IV Utilisation

La détection de pose de MediaPipe permet de suivre les mouvements humains en temps réel. Elle fournit des points de repère pour différentes parties du corps, tels que le nez, les épaules, les coudes, les poignets, les hanches, les genoux et les chevilles. [3]

.IV.1 Reconnaissance faciale

.IV.1.a Définition

Le module de détection de visage analyse une image ou une vidéo pour identifier les régions potentielles où se trouvent des visages. Il utilise des modèles d'apprentissage profond pour effectuer cette détection avec précision. [5]

.IV.1.b Localisation des Points Clés

Une fois qu'un visage est détecté, le module de reconnaissance faciale localise les points clés importants tels que les yeux, les sourcils, le nez et la bouche.

Ces points clés sont essentiels pour caractériser la structure du visage.

.IV.1.c Extraction de Caractéristiques

Les coordonnées des points clés et d'autres caractéristiques du visage sont extraites, fournissant une représentation numérique unique pour chaque visage. Ces représentations permettent de distinguer un visage d'un autre.

.IV.1.d Comparaison ou Correspondance

Les représentations numériques obtenues sont comparées à celles stockées dans une base de données pour identifier un visage spécifique. Cette étape peut également être utilisée pour vérifier l'identité d'une personne.

.IV.1.e Rétroaction en Temps Réel

En temps réel, les résultats de la détection et de la reconnaissance sont affichés, permettant une interaction continue avec l'utilisateur ou l'intégration dans des applications en temps réel.

.IV.2 Suivi de la main

Le suivi de main avec MediaPipe implique l'utilisation du module spécifique appelé "Hands". Ce module utilise des modèles d'apprentissage profond pour détecter et suivre les mouvements des mains en temps réel. [2]

.IV.2.a Initialisation du Module

Pour commencer, vous devez initialiser le module de suivi de main dans votre code en utilisant la bibliothèque MediaPipe. [2]

.IV.2.b Capture de l'Image ou de la Vidéo

Vous devez fournir une image ou une séquence vidéo qui contient des mains que vous souhaitez suivre.

.IV.2.c Application du Module

Appliquez le module de suivi de main à l'image ou à la vidéo. Ceci renverra les résultats de détection, y compris les coordonnées points clés de la main. [2]

.IV.2.d Analyse des Résultats

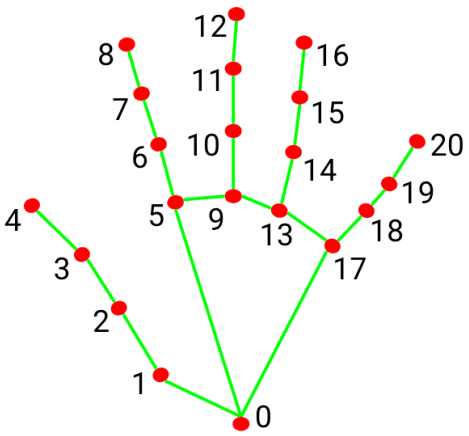
Analysez les résultats renvoyés par le module pour obtenir des informations sur la positionne les mouvements de la main. Les coordonnées des points clés, tel que les bouts des doigts et la base de la paume, peuvent être utilisées pour déterminer la pose de la main.

.IV.2.e Utilisation des Informations de Suivi

Les informations de suivi de main peuvent être utilisées dans diverses applications, telles que des interfaces sans contact, des interactions en réalité virtuelle, des jeux, et d'autres applications interactives.

.IV.2.f Rétroaction en Temps Réel

En temps réel, les résultats du suivi de main peuvent être utilisés pour fournir une rétroaction visuelle ou pour contrôler dynamiquement des éléments dans une application.



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Figure 1: Handlands exemple

.IV.3 Détection d'objet (Objectron module):

Pour la détection d'objets MediaPipe utilise le module Objectron qui se concentre sur la détection d'objets en trois dimensions (3D) en temps réel.Ce module est spécifiquement conçu pour détecter et suivre des objets spécifiques dans des scènes en fournissant des informations sur leur position et leur orientation dans l'espace 3D.

.IV.3.a Modèle d'Apprentissage Profond

Objectron utilise des modèles d'apprentissage profond basés sur des architectures de réseaux neuronaux convolutionnels (CNN) pour effectuer la détection d'objets en 3D.

Ces modèles ont été pré-entraînés sur d'énormes ensembles de données contenant divers objets dans différentes situations et positions.

.IV.3.b Détection en 3D

Lorsque vous appliquez le module Objectron à une image ou à une séquence vidéo, le modèle d'apprentissage profond analyse la scène pour détecter la présence et la position d'objets spécifiques en 3D. Les informations de sortie du modèle incluent les coordonnées spatiales des coins ou des points clés de l'objet détecté.

.IV.3.c Localisation Précise

Objectron vise à fournir une localisation précise des objets dans l'espace 3D, ce qui signifie qu'il fournit des informations sur la position exacte de chaque coin de l'objet détecté par rapport à la caméra.

.IV.3.d Rétroaction en Temps Réel

Les résultats de la détection d'objet peuvent être utilisés en temps réel pour des applications interactives, telles que la réalité augmentée. Par exemple, vous pourriez ancrer des objets virtuels à des objets réels détectés par Objectron.

.IV.3.e Adaptabilité

Les modèles d'Objectron sont conçus pour être robustes et adaptatifs à différentes conditions, ce qui signifie qu'ils peuvent gérer des environnements variés, des angles de vue différents et des objets de différentes tailles.

.IV.3.f Segmentation sémantique

Pour la segmentation sémantique, MediaPipe propose le module appelé Selfie Segmentation. Ce module utilise des modèles d'apprentissage profond pour classer chaque pixel d'une image dans une catégorie spécifique, permettant ainsi de séparer le premier plan (par exemple, une personne) du fond.

.IV.4 Modèle d'Apprentissage Profond

Le module Selfie Segmentation utilise des modèles d'apprentissage profond, souvent basés sur des réseaux neuronaux convolutionnels (Convolutional Neural Network), qui ont été pré-entraînés sur des ensembles de données contenant des images annotées avec des informations de segmentation sémantique.

.IV.4.a Analyse Pixel par Pixel

Lorsqu'on applique le module à une image, le modèle analyse chaque pixel de l'image et le classe dans une catégorie correspondant au premier plan (personne) ou au fond.

.IV.4.b Information de Segment Précis

Le module Selfie Segmentation fournit des informations précises sur les régions de l'image qui appartiennent au sujet (personne) et celles qui appartiennent au fond. Cela permet de créer une segmentation sémantique détaillée de l'image.

.IV.4.c Transparence Dynamique

L'une des caractéristiques intéressantes du module Selfie Segmentation est qu'il peut également fournir une transparence dynamique, permettant d'intégrer le sujet segmenté dans différents environnements virtuels.

.IV.4.d Applications

Les applications de la segmentation sémantique sont variées, allant de l'application de filtres en temps réel sur des visages dans des applications de photographie à l'intégration transparente de personnes dans des arrière-plans virtuels lors de vidéos en direct ou de visioconférences.

.V Discussion

En intégrant de manière experte des réseaux neuronaux profonds, MediaPipe de Google émerge comme un leader visionnaire dans le domaine de la vision par ordinateur.

Ces modules emblématiques tels que la Reconnaissance Faciale, le Suivi de Main, la Détection d'Objet et la Segmentation Sémantique illustrent la maîtrise de MediaPipe dans l'exploitation de l'apprentissage profond. Cette approche novatrice redéfinit notre relation avec la technologie en offrant une compréhension visuelle avancée, ouvrant ainsi la voie à des expériences interactives et intuitives.

Chapitre 3 : Application et Résultats

Chapitre 3 : Application et Résultats

.I Préambule

Le développement d'une application ou d'un système informatique requière l'usage d'une méthodologie afin d'assurer une organisation consciencieuse et de pouvoir cerner les tâches à accomplir.

Figure 2: Logo le l'application

.II Diagramme global

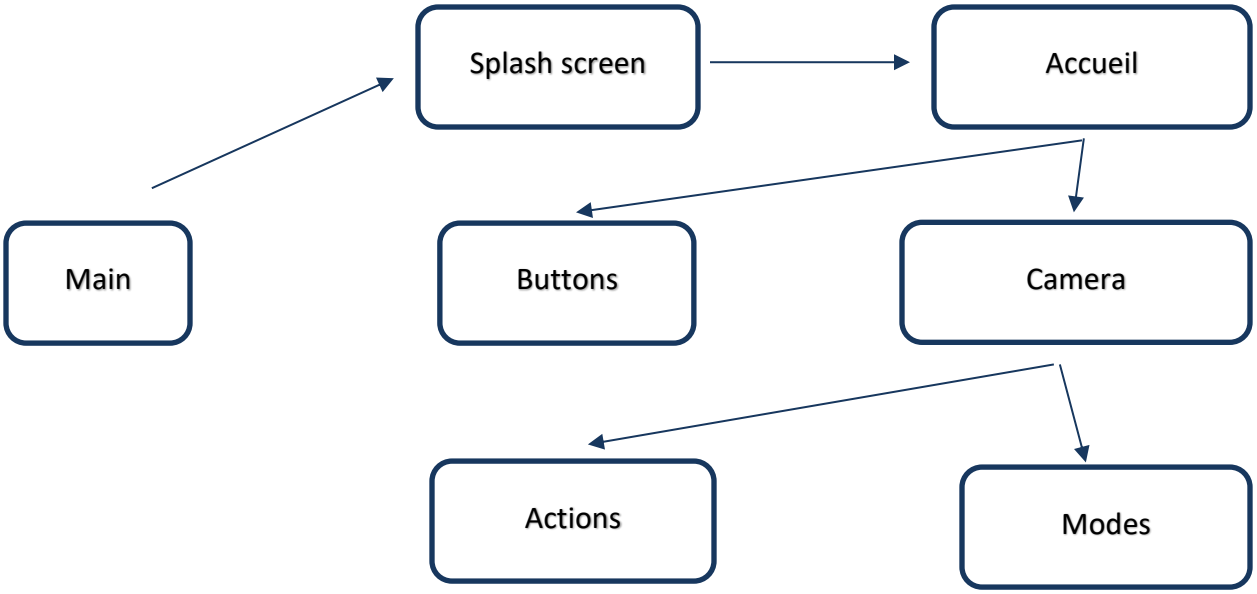


Figure 3: diagramme global de fonctionnement Tâna

.III Interface de l’application

.III.1 Interface principale

L'application Tana Vision vise à capturer des vidéos en temps réel pour la détection et la reconnaissance de gestes de la main. Son interface graphique est conçue avec PyQt6, une bibliothèque utilisant l'API graphique Qt, un framework multiplateforme pour le développement d'interfaces riches et flexibles. PyQt6, en tant que liaison Python pour Qt, permet aux développeurs Python d'exploiter les fonctionnalités de Qt dans la création d'applications avec une interface utilisateur graphique. OpenCV, une bibliothèque spécialisée dans le traitement d'images et de vidéos, est également utilisée. De plus, PyAutoGUI est intégrée pour le contrôle de la souris, offrant des fonctionnalités telles que le déplacement du curseur et les clics. Pour le modèle de la main, MediaPipe est employée, proposant des modules pré-entraînés pour la reconnaissance de gestes.

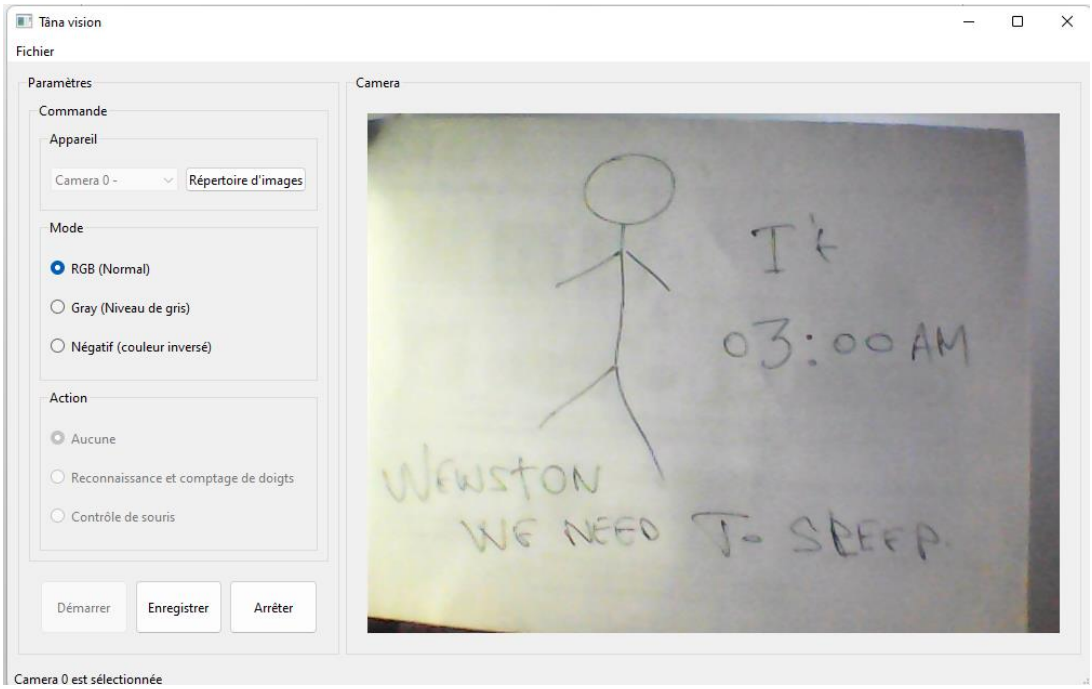


Figure 4: interface principale

.III.2 Espace appareil

Menu déroulant pour le choix de la caméra à utiliser lors de l'acquisition des images.

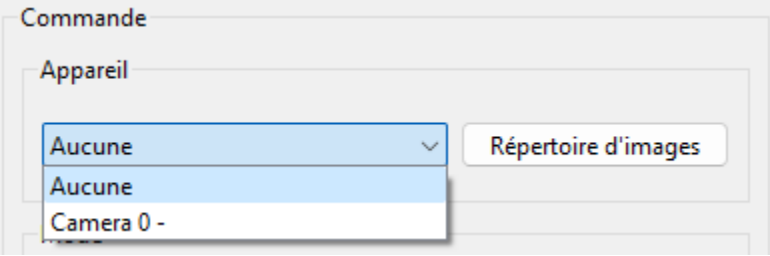


Figure 5: Choix de camera à utiliser

.III.3 Espace Mode

Notre application permet 3 mode d'acquisition et ceci grâce à cette espace mode (RGB, Niveau de gris, …).

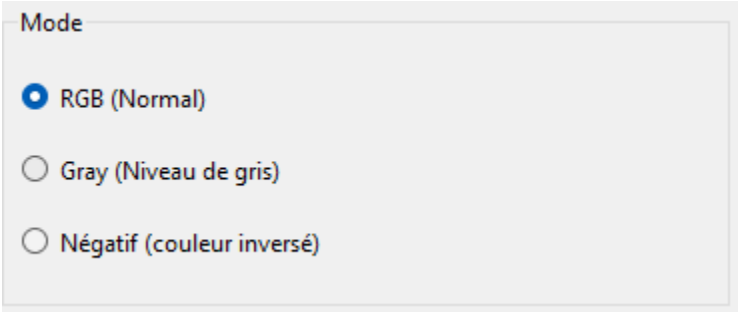
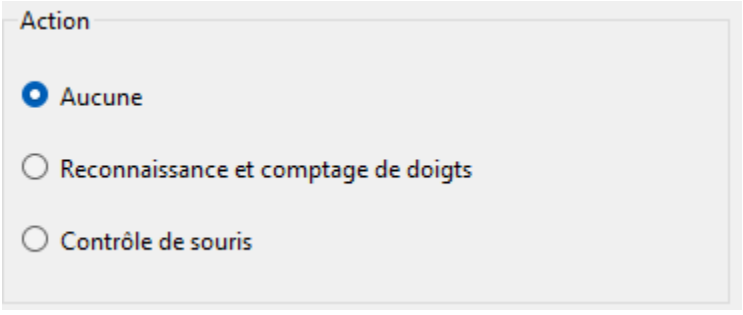


Figure 6 : Choix du mode de la caméra

.III.4 Espace Action

Cet espace nous permet de choisir parmi 3 actions



Ce menu nous permet de démarrer et arrêter la camera ainsi que l'enregistrement des images acquises en temps réel.

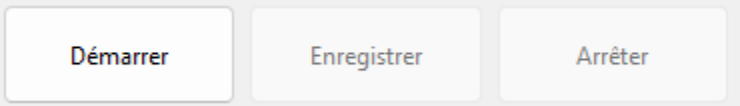


Figure 7 : Boutons de commandes

Figure 8: Espace camera

.III.5 Actions

.III.6 Détection de nombre de doigts

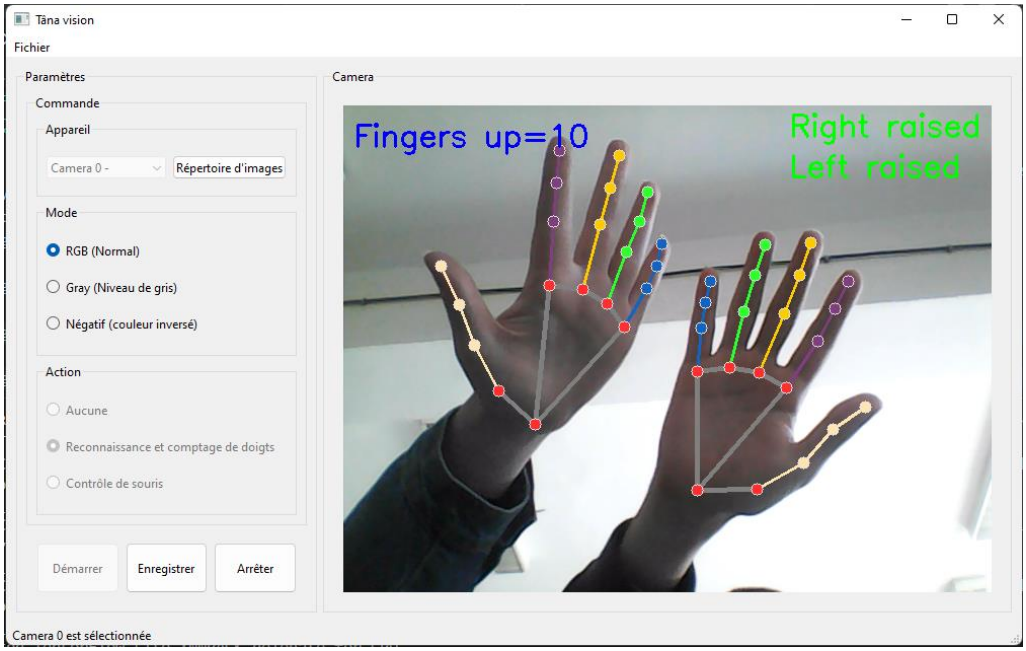


Figure 9 : Exemple détection avec 10 doigts

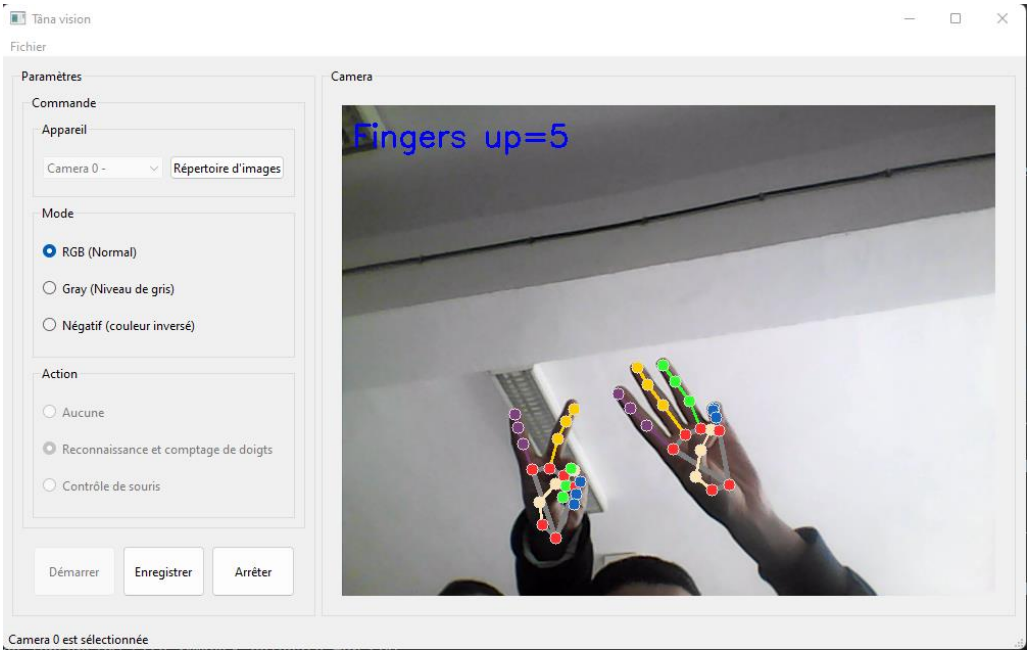


Figure 10 : Exemple détection avec 5 doigts

.III.7 Controller la position souris

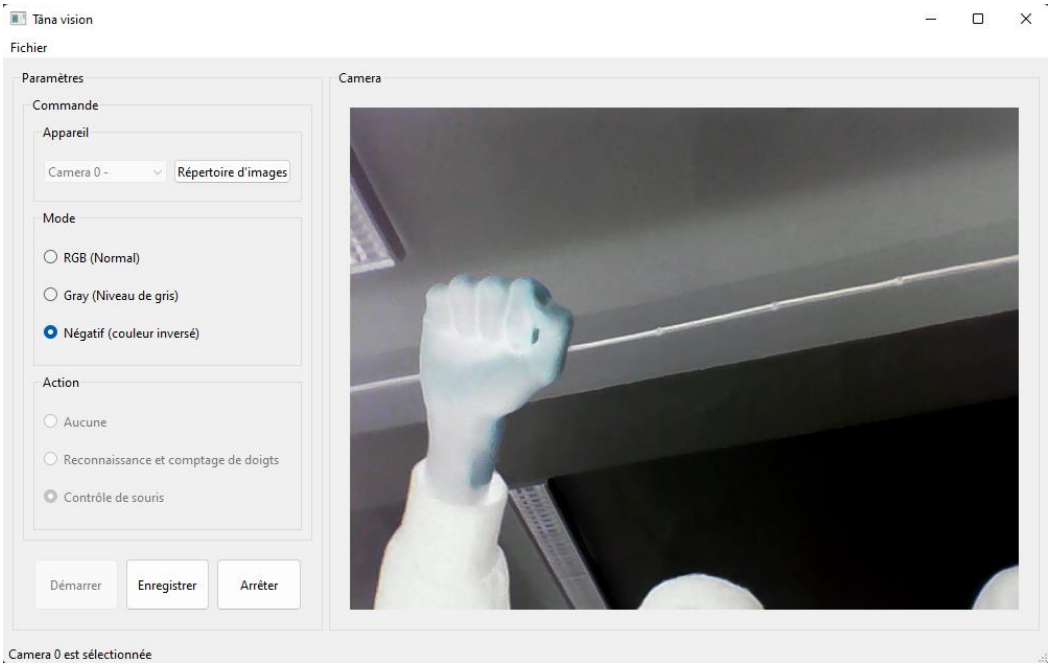


Figure 11 : Exemple de positionnement de la souris avec la main (Mode inversé)

.IV Conclusion

La détection du nombre de doigts et le contrôle de la souris à l'aide de gestes de la main offrent des possibilités prometteuses pour améliorer l'interaction homme-machine. En utilisant la détection de landmarks fournie par des bibliothèques comme MediaPipe, il devient possible de détecter avec précision le nombre de doigts levés et de mapper ces gestes à des actions de la souris. Ces technologies ouvrent la voie à des interfaces utilisateur sans contact, à des expériences immersives en réalité virtuelle, et peuvent également bénéficier aux personnes ayant des limitations physiques en offrant des méthodes alternatives de contrôle de l'ordinateur. Cependant, pour atteindre leur plein potentiel, ces systèmes doivent surmonter des défis tels que la précision de la détection dans différentes conditions et la conception de gestes intuitifs et fiables. En résumé, la fusion de la détection de gestes de la main et du contrôle de la souris représente une direction passionnante pour l'innovation dans l'interaction informatique

Bibliographie

Références

[1] G. Garcia-Hernando, S. Yuan, S. Baek, and T.-K. Kim, “First-person hand action benchmark with rgb-d videos and 3d hand pose annotations,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 409–419, 2018.

[2] Q. De Smedt, H. Wannous, J.-P. Vandeborre, J. Guerry, B. L. Saux, and D. Filliat, “3d hand gesture recognition using a depth and skeletal dataset : Shrec’17 track,” in Proceedings of the Workshop on 3D Object Retrieval, 3Dor ’17, (Goslar, DEU), p. 33–38, Eurographics Association, 2017.

[3] F. M. Caputo, S. Burato, G. Pavan, T. Voillemin, H. Wannous, J.-P. Vandeborre, M. Maghoumi, E. Taranta, A. Razmjoo, J. LaViola Jr, et al., “Online gesture recognition,” in Eurographics Workshop on 3D Object Retrieval, The Eurographics Association, 2019.

[4] T. Voillemin, H. Wannous, and J.-P. Vandeborre, “2d deep video capsule network with temporal shift for action recognition,” in 2020 25th International Conference on Pattern Recognition (ICPR), pp. 3513–3519, IEEE, 2021

[5] T. Voillemin, H. Wannous, and J.-P. Vandeborre, “Deep video capsule network avec décalage temporel pour la reconnaissance d’action,” in admis à COmpression et REprésentation des Signaux Audiovisuels (CORESA), CORESA, 2021.