

Controlador de jocs clàssics amb gestos de mà.

Marçal Armengol Romero (1668314), Pere Llauradó Adeva (1669502)

Abstract— Aquest projecte presenta un sistema de control de videojocs clàssics com Snake o Pong (extrets de GitHub) a través de gestos de mà detectats per la càmera d'un ordinador. Ens basarem en aconseguir un aplicació robusta del reconeixement de la gestualitat de la mà més que en la implementació dels jocs. Utilitzant tècniques de visió per computador i llibreries com Mediapipe i biblioteques com OpenCV i Pygame, es proposa una arquitectura que permet interpretar moviments simples de la mà per controlar elements del joc. L'objectiu és oferir una experiència d'interacció natural, accessible i educativa per a usuaris de qualsevol nivell tècnic.

Keywords—Visió per computador, reconeixement de gestos, control per càmera, segmentació per colors, blurring, erode, dilate

1 INTRODUCCIÓ

La interacció sense contacte està guanyant protagonisme en diversos àmbits com la robòtica, els videojocs o les aplicacions mèdiques. Aprofitant aquest avenç, el nostre projecte proposa controlar jocs clàssics com el *Snake* o el *Pong* amb moviments de mà captats per la càmera web. Aquest enfocament busca no només oferir una alternativa divertida a les interfícies tradicionals, sinó també fomentar l'ús de la visió per computador en entorns educatius i accessibles. Usarem estratègies apreses de visió per computador per intentar obtenir una robustesa més gran del nostre detector. Farem proves gravant-nos a nosaltres mateixos per la càmera web. Buscarem fer proves amb diferents tipus d'il·luminació, amb qualitat baixa d'imatge o fent que la mà no estigui en el pla principal. També provarem com de robust és el nostre treball quan hi ha molt moviment. Tant en el codi utilitzant Media Pipe com utilitzant tècniques clàssiques de visió per computador.

Els objectius principals són:

- Desenvolupar un sistema de detecció de gestos de mà en temps real.
- Millorar fins al màxim possible la robustesa del nostre projecte per a poder detectar la mà en situacions més complicades
- Implementar la lògica de control d'un joc senzill que respongui als gestos.
- Avaluar la precisió i la usabilitat del sistema.

2 ESTAT DE L'ART

2.1 Introducció

El reconeixement de gestos de la mà ha estat àmpliament estudiat dins el camp de la visió per computador, amb aplicacions que van des del control d'interfícies fins al reconeixement del llenguatge de signes. Els sistemes existents es poden agrupar en dues grans categories: aquells que utilitzen biblioteques avançades i models entrenats, i aquells que implementen mètodes clàssics basats en processament d'imatge.

Pel que fa als mètodes clàssics, aquests aprofiten tècniques com la segmentació per color (per exemple, mitjançant transformacions a l'espai HSV i llindars adaptats al to de pell), la detecció de contorns per extreure la silueta de la mà, i l'ús de punts de forma o convex hull per identificar dits o gestos específics. També són comuns les tècniques de background subtraction per aïllar la mà del fons, i l'ús d'histogrames de gradients (HOG) o descriptors similars per a la classificació de formes. Aquest enfocament permet una millor comprensió dels fonaments de la visió per computador i una aplicació directa dels conceptes teòrics apresos.

En el context educatiu, molts projectes opten per aquest tipus d'implementació per reforçar l'aprenentatge pràctic. A més, aquests mètodes tenen l'avantatge de ser lleugers i comprensibles, tot i que solen ser més sensibles a condicions com la llum, el fons o la variabilitat de les mans. En el nostre cas ens interessa poder fer alguna versió simple per tenir una comprensió general del funcionament d'aquesta i intentar millorar-la.

Pel que fa al control de jocs, diversos projectes han integrat detecció de gestos per controlar videojocs clàssics. Per exemple, en el cas de *Snake* o *Pong*, s'han proposat sistemes on es detecta el moviment horitzontal o vertical de la mà per moure el personatge o la pala, respectivament. Alguns d'aquests projectes combinen la detecció de moviment amb classificació bàsica per identificar gestos com “mà oberta”, “puny tancat” o “un dit aixecat”. Aquests sistemes demostren que, fins i tot sense xarxes neuronals ni models preentrenats, és possible crear una interacció fluida i efectiva.

Cal destacar que mentre les biblioteques com *MediaPipe* ofereixen solucions molt robustes i fàcils d'integrar, el desenvolupament d'un sistema de detecció propi permet entendre millor els reptes i les limitacions del reconeixement visual en temps real. Aquest projecte aposta per aquesta segona via, afavorint l'experimentació amb tècniques apreses a classe per construir una interfície de control natural per videojocs prefets.

Finalment, encara que usarem llibreries ja fetes de reconeixement com *MediaPipe* o *Yolov5* voldrem no abusar-ne i intentar dissenyar la nostra pròpia algorítmica per fer la detecció de mans. Intentarem que les llibreries ens serveixin com un punt de suport en el que basar-nos per començar amb el projecte.

2.2 Exemple d'un projecte

Un exemple rellevant és el projecte “Hand Gesture Controlled Ping Pong” [1], desenvolupat amb Python i OpenCV. És un projecte realitzat en un repositori de GitHub per l'usuari “EmrNITK” que intenta assolir alguns objectius semblants als nostres. Aquest projecte permet controlar una pala de ping-pong mitjançant gestos de mà detectats per la càmera web. Utilitza tècniques de segmentació d'imatge i detecció de contorns per identificar la posició de la mà en temps real, traduint els moviments horitzontals en desplaçaments de la pala dins del joc. Aquest enfocament demostra la viabilitat d'utilitzar mètodes clàssics de visió per computador per a la interacció amb videojocs sense necessitat de maquinari especialitzat. També usarem aquest treball i algunes de les seves funcions per poder tenir un punt de partida pel nostre projecte.

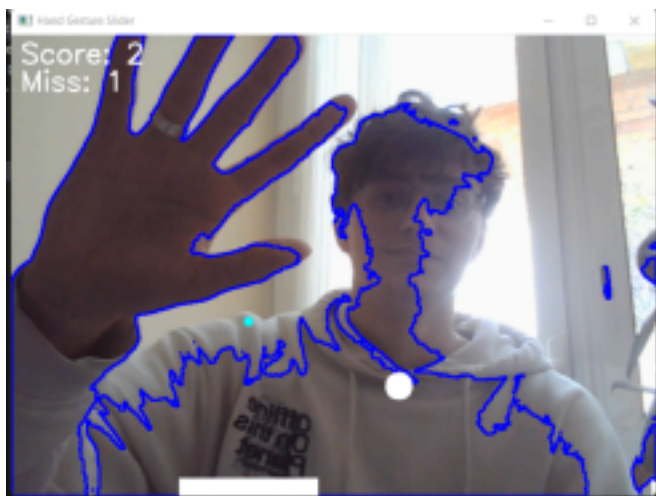


Fig. 1. Prova d'una proposta amb objectius similars als nostres ja realitzada en un repositori del GitHub.

En la *fig. 1* podem veure quins són els resultats ja disponibles en repositoris accessibles per a tothom. En aquest exemple, el projecte es basa en el reconeixement de tot el cos i en com movem la mà per moure la paleta. Això ens serveix per entendre com funciona però la nostra idea és diferent. Nosaltres també volem reconèixer tot el cos però ens centrarem únicament en el moviment i les diferents posicions que pot prendre la mà.

3 DESCRIPCIÓ DE LES DADES

Per entrenar i validar els algoritmes de detecció de mans i reconeixement de gestos, el projecte farà ús tant de dades pròpies com de datasets públics disponibles en plataformes com Kaggle.

D'una banda, es recolliran vídeos propis enregistrats amb la càmera web de l'ordinador, on els membres del grup realitzaran els gestos que s'utilitzaran al sistema (per exemple: mà oberta, puny tancat, moviment cap a l'esquerra o dreta, dit índex estès, etc.). Aquests vídeos es gravaran en diverses

condicions de llum i amb diferents fons per augmentar la variabilitat i simular escenaris reals. A partir d'aquests enregistraments, s'extreurà

una sèrie de fotogrames que es faran servir per generar les mostres d'entrenament i test, aplicant-hi segmentació, etiquetatge manual i augmentació de dades (rotacions, zoom, soroll, etc.) si cal.

D'altra banda, es farà ús de datasets públics com ara:

- *Hand Gesture Recognition Database* (Kaggle): amb imatges etiquetades de gestos estàtics en diferents posicions i angles.
- *EgoHands Dataset*: encara que enfocat a detecció de mans en entorns naturals, pot servir com a suport per entrenar models de segmentació.

Aquest conjunt de dades mixt, que combina vídeos personalitzats amb bases de dades reals, permetrà entrenar models més robustos i generalitzables, i facilita tant la fase de desenvolupament com la validació del sistema.

4 EXPERIMENTS, RESULTATS I ANÀLISI

El problema que es planteja és el de reconèixer gestos de mà a partir d'imatges captades en temps real amb una càmera web, amb l'objectiu de controlar videojocs clàssics. Per resoldre aquesta tasca, s'ha optat per implementar tècniques de visió per computador clàssiques apreses a classe i per altra banda utilitzarem biblioteques avançades com *MediaPipe* per tal de comparar els dos mètodes i fomentar la comprensió i l'aprenentatge actiu.

El desenvolupament del programa es dividirà en dues parts:

1. La primera buscant aplicar el model de detecció de mans de *MediaPipe* i aplicar-lo al joc del snake, per a veure el funcionament tan robust que assoleix aquesta llibreria.
2. La segona, consistirà en aplicar mètodes vist a classe de visió per computador per a poder mostrar com es pot assolir un resultat mínimament robust, encara que sigui en unes condicions determinades, per a detectar la mà i la direcció.

Aquest enfocament permet resoldre el problema de forma modular i comprensible, amb tècniques totalment fonamentades en els continguts vistos a classe. A més, permet una futura extensió cap a tècniques més avançades si es volgués millorar la robustesa o reconèixer gestos més complexos. Per altra banda, comparar dos mètodes, un més complex i de difícil comprensió creat per professionals i un altre creat de zero per nosaltres, d'aquesta manera podrem veure la diferència de funcionament.

4.1 Jocs a implementar el detector

Per tal de validar el sistema de reconeixement de gestos desenvolupat, s'integrarà en dos videojocs clàssics: Pong i Snake, extrets de repositoris públics de GitHub i adaptats per acceptar entrada visual.

En el cas de Pong, es tracta d'una versió per a un sol jugador on la pala s'ha de moure verticalment per retornar la pilota. El moviment de la pala es controlarà detectant el desplaçament horitzontal de la mà davant de la càmera o amb gestos indicant esquerra o dreta. [2]

Pel que fa al Snake, el jugador haurà de guiar la serp perquè reculli objectes sense xocar amb les parets ni amb ella mateixa. En la *figura 2* podem veure'n un exemple modificable extret de GitHub [3]. Els moviments de la mà en quatre direccions (esquerra, dreta, amunt, avall) es maparan a les tecles de direcció del joc.

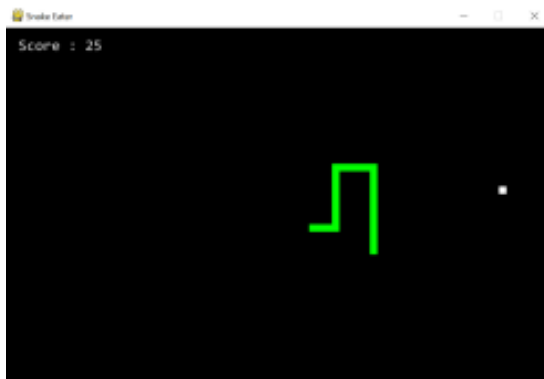


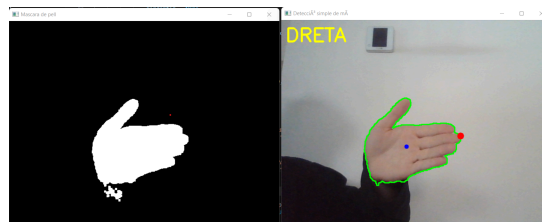
Fig. 2. Implementació senzilla del joc *snake* que podem usar pel testing del nostre projecte.

Aquests dos jocs permeten provar gestos tant estàtics com dinàmics, i representen escenaris ideals per avaluar la precisió i la latència del sistema de detecció.

4.2 Detector posició de la mà propi

Com hem explicat anteriorment en aquesta part del projecte hem buscat una solució senzilla per detectar i interpretar el moviment d'una mà captada per la càmera, utilitzant tècniques bàsiques de visió per computador implementades amb OpenCV i NumPy. El sistema no requereix l'ús de models d'aprenentatge automàtic ni biblioteques avançades com MediaPipe, sinó que es basa en processament clàssic d'imatge i anàlisi geomètrica de contorns.

Aquest codi funciona com un sistema bàsic de reconeixement gestual en temps real. Primer, accedeix a la càmera i captura imatges continuament. A continuació, identifica les zones de la pell a la imatge aplicant una transformació de color i una filtració per reduir el soroll aplicant una erosió i una dilatació. Un cop detectada la mà, n'extreu el contorn principal i calcula el seu centre. A partir d'aquí, analitza la posició del dit més estès en relació amb aquest centre per determinar cap a on apunta la mà: amunt, avall, dreta o esquerra. Finalment, mostra per pantalla tant la direcció detectada com una visualització de la detecció mostrant per pantalla la direcció cap on apunta la mà detectada en temps real com es pot veure a la *Fig. 3*.



El resultat final, com he dit anteriorment és una interfície que detecta la mà i interpreta gestos bàsics de direcció (cap a quina direcció apunta el dit més destacat). Tot i així, és molt menys robust el sistema que el que hem utilitzat de la llibreria MediaPipe, ja que s'ha de procurar tenir un fons de color uniforme i que només es vegi la mà a la càmera, tapant, si pot ser, el canell i el braç amb una peça de roba de color fosc per a que pugui detectar amb un gran encert la posició de la mà.

4.3 Detector posició de la mà amb mediapipe

Pel que fa als resultats aquests els hem dividit en 3 fitxers. El primer, `deteccioMansmediaPipe` on, usant la llibreria [4] *mediapipe* hem implementat un codi que permet reconèixer punt d'interès de la mà (llibreria ja entrenada per fer-ho) i que permet detectar en quina posició (entre esquerra, dreta, amunt i avall) està la mà en tot moment. Abans d'usar les funcions donades per la llibreria, però, s'ha fet preprocessament de la imatge per poder-la usar.

Per començar, es redueix a la meitat el tamany dels canals per evitar que sigui una imatge massa pesada i que no puguem fer la detecció a una velocitat alta (l'objectiu és poder fer un joc interactiu). També utilitzem un filtre gaussià per suavitzar la imatge de manera ràpida i evitar que hi hagi soroll en la imatge. Finalment, passem la imatge a l'espai de color RGB que ens permet usar totes les funcions correctament de la llibreria.

Com podem veure en la *figura 4* el resultat obtingut és un detector de posició de la mà que és força robust. Detecta qualsevol color de pell, no és sensible a la il·luminació i permet detectar la direcció de forma quasi immediata. També s'ha fet una petita interfície en forma de finestra per poder visualitzar quin és el funcionament i resultat d'aquest mòdul.

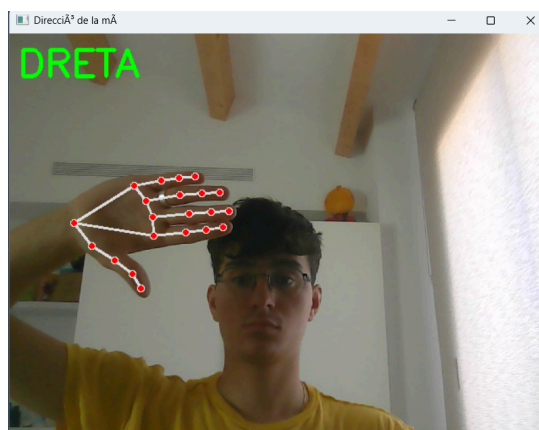


Fig 4.

Detecció de les mans que hem dut a terme en el nostre projecte usant la llibreria *mediapipe*

En la imatge podem veure com la llibreria agafa punts d'interès de la nostre mà (com les falanges dels dits i quin és el punt més allunyat d'aquests) per poder-ne detectar quina és la posició que l'usuari vol indicar. També observem com, tot i que la il·luminació no és molt bona, el reconeixement segueix funcionant sense cap mena de dificultat.

4.4 Implementació de Snake

Tot i que aquest no era l'objectiu principal del nostre projecte, una part que també hem desenvolupat ha estat l'integració del detector de moviments de la mà realitzat amb un joc senzill com l'Snake. Aquest l'hem implementat nosaltres mateixos i hem obtingut una interfície senzilla però efectiva que ens permet interaccionar amb el detector de mans.

La llibreria que ens ha permès dur a terme aquesta implementació ha estat pygame [5].

Com podem veure en la figura 5 tenim una interfície de joc senzilla però que ens ajuda a tenir clar com podem usar més endavant alguna aplicació com la que hem fet.

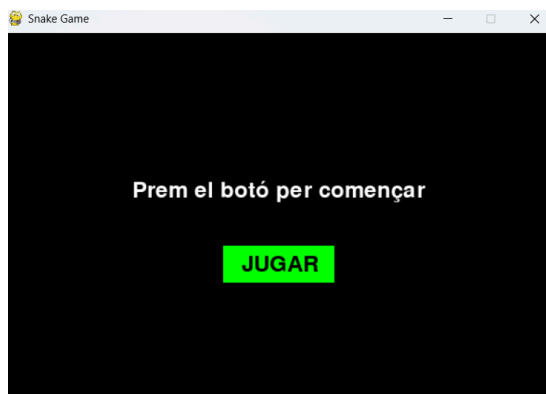


Fig. 5. Implemetació senzilla del clàssic joc Snake usada pel nostre projecte

5 CONCLUSIONS

Al llarg d'aquest projecte hem desenvolupat dues solucions diferents per detectar la direcció cap a on apuntem amb la mà. D'una banda, em creat un sistema basat en tècniques clàssiques de visió per computador, i de l'altra hem fet ús de la llibreria MediaPipe, entrenada específicament per al reconeixement de mans. Aquesta segona, també la hem pogut connectar al joc clàssic del Snake on es podia controlar els moviments de la serp a través dels gestos de la mà sense cap mena de problema.

La principal conclusió que en podem extreure és que, si bé ambdós enfocaments són funcionals, la robustesa del sistema basat en MediaPipe és molt superior. Aquest sistema permet detectar mans de qualsevol color de pell, amb diferents condicions d'il·luminació i en escenaris amb moviment, soroll visual o distàncies diferents, mantenint una resposta ràpida i precisa. A més, requereix molt poc ajust manual i ofereix una

detecció consistent i que no requereix de molts recursos del ordinador per funcionar, independentment del fons o la roba de l'usuari. Per tant gràcies a la seva robustesa i eficiència l'hem pogut aplicar al joc comentat anteriorment sense que es produeixin greus problemes pel lag.

En canvi, el sistema basat en algorismes clàssics funciona correctament, però és molt més sensible a les condicions de l'entorn. Per aconseguir bons resultats, cal garantir un fons neutre (que no sigui del mateix color de la mà) i portar màniga llarga fosca per tal de reduir interferències visuals amb el canell i el braç. Tot i aquestes limitacions, aquest mètode ens ha permès entendre profundament els fonaments del processament d'imatge i ha estat molt útil des d'un punt de vista educatiu. A part les dues pantalles que es mostren permeten veure que fa el algorisme per a detectar els diferents blobs i l'altra mostra el resultat dels càlculs per a saber la direcció on apunta la mà.

Finalment, remarcuem que l'ús d'un sistema clàssic ens ha permès construir una solució pròpia i modular, fàcilment adaptable i comprensible, mentre que MediaPipe ha servit com a referència de qualitat i fiabilitat per a sistemes reals. Per tant, veure per una banda sistemes molt precisos i complexos i crear el nostre sistema a partir del que hem après a classe, ens ha permet obtenir una visió sobre el gran potencial que té el que estem aprenent amb aquesta assignatura i tot el que amb els coneixements que tenim actualment podem fer, d'una manera molt visual i entretinguda.

BIBLIOGRAFIA

- [1] Hand gesture controlled ping pong - GitHub:
<https://github.com/EmrNITK/Ping-Pong>
- [2] Pong python - GitHub:
<https://github.com/skar91/pong-python>
- [3] Snake eater - GitHub:
<https://github.com/rajatdiptabiswas/snake-pygame>
- [4] MediaPipe-Google
<https://ai.google.dev/edge/mediapipe/solutions/guide?hl=es-419>
- [5] Pygame
<https://www.pygame.org/news>