

## Interfície *head-tracking* per entorns mòbils

Miquel Àngel Alcalde Velado  
Grau en Enginyeria Informàtica  
05.629 TFG - Intel·ligència artificial

**Dr. Joan M. Nunez do Rio**  
**Dr. Carles Ventura Royo**

02/05/2019



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	<i>Interfície head-tracking per entorns mòbils</i>
<b>Nom de l'autor:</b>	<i>Miquel Àngel Alcalde Velado</i>
<b>Nom del consultor/a:</b>	<i>Dr. Joan M. Nunez do Rio</i>
<b>Nom del PRA:</b>	<i>Dr. Carles Ventura Royo</i>
<b>Data de lliurament (mm/aaaa):</b>	<i>05/2019</i>
<b>Titulació o programa:</b>	<i>Grau en Enginyeria Informàtica</i>
<b>Àrea del Treball Final:</b>	<i>05.629 TFG - Intel·ligència artificial</i>
<b>Idioma del treball:</b>	<i>Català</i>
<b>Paraules clau</b>	<i>Computer vision, head-tracking, Human-computer interaction</i>
<b>Resum del Treball:</b>	
<p>Els avenços tecnològics de les darreres dècades ha propiciat l'aparició de noves formes de comunicar-nos amb les màquines. En les interfícies d'interacció natural, la comunicació amb la màquina es realitza sense necessitat de dispositius auxiliars, sinó a través de les capacitats naturals de l'individu.</p> <p>El projecte implementa una interfície natural per detecció de moviments del cap per a dispositius mòbils. La divisió modular dels components en capes de responsabilitat facilita que el sistema sigui portat amb menor esforç a noves plataformes.</p> <p>Per a aquest treball la capa dependent de la plataforma s'ha desenvolupat per a dispositius Android, mentre que la capa corresponent a l'algorisme de visió de la solució ha estat desenvolupador en llenguatge C++ i les llibreries de visió OpenCV.</p> <p>L'algorisme de visió implementat processa el flux d'imatges proporcionades per la càmera del dispositiu, i executa els components de algorisme de visió, per a determinar la posició en pantalla del punter virtual associat d'un punt imaginari, situat en el centre de la punta del nas de l'usuari.</p> <p>Aquests components porten a terme tasques de detecció de seguiment del punter virtual, detecció del gest de selecció per part de l'usuari o la detecció de les condicions d'il·luminació.</p> <p>S'ha desenvolupat una aplicació per generar mètriques i indicadors de funcionament que permeten validar el comportament dels components. Fent de la solució una llavor per futures millores, tant en els algorismes de visió com en el tema de la plataforma.</p>	
<b>Abstract:</b>	

Technological advances in recent decades has led development of new ways to communicate with the machines. On user natural interaction interfaces, the communication human-machine is done without auxiliary devices, but through the natural capacities of the user.

The project implements a natural user interface based on the detection of head movement for mobile devices. The modular division of the components in layers of responsibility ease the system to be ported with less effort to new platforms.

In this project, the platform-dependent layer has been developed for Android devices, while the layer corresponding to vision algorithm has been a developer in C ++ language and OpenCV vision framework.

The implemented vision algorithm processes the flow of images provided by the device's camera, and execute vision components, to determine the on-screen position of the pointer associated of the imaginary point, located in the center of of the user nose.

Components of vision algorithm carry out several task as tracking detection tasks of the virtual pointer, detection of the selection gesture by the user or the detection of lighting conditions.

An application has been developed to generate metrics and indicators that allow validating the behavior of the components. Making the solution a seed for future improvements, both in vision algorithms and in the subject of the platform.

# Índex

1. Introducció.....	1
1.1 Context i justificació del Treball.....	1
1.2 Objectius del Treball.....	4
1.3 Enfocament i mètode seguit.....	6
1.4 Planificació del Treball.....	7
1.5 Fites principals.....	8
1.6 Breu sumari de productes obtinguts.....	8
1.7 Estructura del document.....	9
2. Interfícies de visió per computador.....	10
2.1. Visió per computador.....	10
2.2. Reconeixement de patrons.....	11
2.2. Interfícies <i>head-tracking</i> .....	16
3. Disseny del sistema.....	20
3.1. Component Widget.....	21
3.2. Component HeadTrackingBridge.....	22
3.3 Component HeadTracking.....	22
3.4. Component HeadTrackingClick.....	22
3.4. Component Exposition.....	23
4. Algorisme de visió.....	26
4.1. Fase de detecció.....	27
4.2. Fase de seguiment.....	32
4.3. Fase d'estimació.....	33
4.4. Fase de transferència.....	34
4.5. Estabilitat del procés del seguiment.....	37
4.6 Detecció de l'esdeveniment de <i>click</i> .....	41
4.7 Funcionalitats addicionals.....	45
5. Aplicació de verificació.....	49
5.1 Anàlisi dels resultats d'indicadors i mètriques.....	51
6. Conclusions.....	56
7. Glossari.....	58
8. Bibliografia.....	60

## Llista de figures

Figura 1: Components del sistema.....	6
Figura 2: espai de color RGB.....	12
Figura 3: espai de color HSV.....	13
Figura 4: espai de color YcbCr.....	13
Figura 5: característiques de Haar.....	15
Figura 6: estructura en cascada.....	16
Figura 7: fases interfície <i>head-tracking</i> genèrica.....	16
Figura 8: relació de fases filtro de Kalman.....	19
Figura 9: components del sistema.....	20
Figura 10: diagrama de classes.....	24
Figura 11: diagrama de seqüència.....	25
Figura 12: entrada i sortida de l'algorisme.....	26
Figura 13: fase de detecció.....	27
Figura 14: fallades de detecció frontal.....	27
Figura 15: error de detecció múltiple.....	28
Figura 16: regles antropomòrfiques del rostre.....	29
Figura 17: regió del nas afitada.....	29
Figura 18: filtratge de punts per distància del punt central.....	31
Figura 19: aparellament de punts.....	31
Figura 20: detecció de punts fora de l'àrea fitada.....	33
Figura 21: punt mitjà i punt mitjà de Kalman.....	34
Figura 22: posició de l'usuari desplaçada del centre.....	35
Figura 23: desplaçament de la càmera.....	35
Figura 24: mapa de calor moviment del nas.....	36
Figura 25: relació espais de coordenades.....	37
Figura 26: esquema de buffer de <i>frames</i> .....	37
Figura 27: màquina d'estats de l'algorisme.....	38
Figura 28: divisió de tasques per estat.....	39
Figura 29: visió general de processos de l'algorisme.....	40
Figura 30: fases del algorisme detecció gesto wink.....	41
Figura 31: processos de l'algorisme detecció gesto wink.....	41
Figura 32: tasques d'obtenció imatge binari dels ulls.....	42
Figura 33: polígons d'aproximació als ulls.....	43
Figura 34: detecció del gesto mitjançant polígons d'aproximació.....	43
Figura 35: fals negatiu del detector d'ulls.....	44
Figura 36: algorisme de click per dwell-time.....	44
Figura 37: <i>frame</i> RGB, HSV, canal S i histograma resultant.....	45
Figura 38: algorisme del valor d'exposició.....	46
Figura 39: categorització del valor màxim d'exposició.....	46
Figura 40: graella de zones de desplaçament.....	47
Figura 41: divisió en zones de cobertura.....	48
Figura 42: graella de posicions de l'aplicació.....	49
Figura 43: imatge binària del cicle de joc.....	51
Figura 44: correlació de l'indicador d'exposició.....	53
Figura 45: correlació de l'indicador de posicionament i cobertura.....	53
Figura 46: correlació de l'indicador desplaçament.....	54
Figura 47: mapa de calor de les sessions dels usuaris.....	54

## Lista de taules

Taula 1: gestió de riscos del projecte.....	7
Taula 2: planificació del treball.....	7
Taula 3: fites del projecte.....	8
Taula 4: opcions de configuració del detector.....	28
Taula 5: opcions de configuració del càlcul de punts característics.....	30
Taula 6: opcions de configuració de l'algoritme de Lukas-Kanade.....	32
Taula 7: opcions de configuració del model del filtre de Kalman.....	34
Taula 8: categorització de les condicions d'il·luminació.....	46
Taula 9: categorització del posicionament de l'usuari.....	47
Taula 10: categorització del grau de cobertura.....	48
Taula 11: indicadors de l'aplicació de verificació.....	50

# 1. Introducció

## 1.1 Context i justificació del Treball

Els avenços tecnològics ocorreguts en les darreres dècades han situat les tecnologies de la informació com un dels elements centrals de les societats i economies del segle XXI. El nivell d'innovació i influència de les tecnologies de la informació continuarà creixent en les pròxima dècades.

En la cerca constant d'innovació, els investigadors i fabricants han treballat conjuntament en el disseny de noves formes de comunicar-nos amb les màquines. Mètodes de comunicar-se amb les màquines que es consideraven ciència-ficció són ara una realitat quotidiana, com és el cas de les interfícies basades en ordres de veu dels nous altaveus intel·ligents o les interfícies tàctils incorporades en els nostres telèfons mòbils.

La interacció de les persones amb les màquines és objecte d'estudi de la disciplina *Human-Computer Interface (HCI)* [1-2]. Aquesta àrea d'estudi és multidisciplinària i inclou aspectes de la informàtica, la psicologia, ciències socials i del disseny industrial.

La disciplina **HCI** es dedica a dissenyar, avaluar i implementar sistemes informàtics interactius per a l'ús humà, i a estudiar els fenòmens relacionats més significatius.

La disciplina **HCI** està relacionada amb altres àrees d'estudi de les interfícies usuari-màquina com són:

- **L'ergonomia:** és la disciplina que s'encarrega del disseny de llocs de treball per a adaptar-los a les característiques anatòmiques, psicològiques i les capacitats dels usuaris.
- **L'usabilitat:** és la disciplina que estudia la facilitat amb què les persones poden utilitzar una eina en particular amb la finalitat d'aconseguir un objectiu concret.
- **L'accessibilitat:** és l'estudi d'aquells aspectes que permeten als usuaris accedir i, utilitzar un objecte o servei independentment de les seves capacitats tècniques, cognitives o físiques.

Malgrat totes les recomanacions i estudis realitzats dins de l'àmbit de **HCI**, és freqüent que els nous paradigmes d'interacció usuari-màquina sorgits en els últims temps no sempre resultin accessibles per a les persones amb discapacitat. Les dificultats en l'ús de les tecnologies de la informació ha generat preocupació a causa de la bretxa digital que suposa i té com a conseqüència directa l'aparició de cert grau de marginalitat [3][4] per aquest col·lectiu.



Segons dades de l'organització mundial de la salut, el 15.6% de la població mundial [5] ,8.6% en el cas d'Espanya [6], presenta algun grau de discapacitat. Per aquest motiu els governs i diverses institucions estan promovent iniciatives d'igualtat tecnològica per les persones discapacitades.

Una interfície multimodal és aquella que permet a l'usuari la interacció a través de diferents canals. Els desenvolupadors poden incorporar diverses maneres d'interactuar amb les aplicacions per així garantir l'accessibilitat. A través de la interfície alternativa les persones amb discapacitat poden tenir funcionalitats anàlogues que els permetin l'ús de les aplicacions de forma igualitària. Un exemple d'aquest tipus d'interfícies són les recomanacions *Web Accessibility Initiative (WAI)* promoguda pel consorci *World Wide Web (W3C)* o les noves interfícies per reconeixement de veu incorporades en els sistemes operatius.

Conseqüència dels treballs de recerca en el desenvolupament de noves interfícies no invasives s'ha desenvolupat el concepte d'interfícies naturals (*Natural User Interface, NUI* [7]). Les interfícies **NUI** són aquelles en les quals la interacció amb l'aplicació es realitza sense utilitzar sistemes de comandament o dispositius d'entrada. En el seu lloc. es fa ús de les capacitats naturals de cos humà com la veu, o el tacte, o moviments gestuals del cos (mans, cap o ulls). Sense dubte l'exemple més conegut és el dispositiu **Xbox Kinect** desenvolupat per **Microsoft**.

Les capacitats de computació dels sistemes actuals, així com els últims avenços en intel·ligència artificial han permès el desenvolupament d'interfícies **NUI** basades en visió cada vegada més potents i robustes.

Els algorismes de visió faciliten la construcció d'interfícies que permeten una interacció més accessible, característica rellevant en el cas de les persones discapacitades [9-13]. Aquest nou tipus d'interfícies no invasives poden ser aplicades amb èxit en el cas de persones amb discapacitades motrius en mans o braços que els impedeixen l'ús d'un altre tipus d'interfícies tradicionals com a teclats i ratolins.

En el cas dels telèfons mòbils intel·ligents, comprovem que, malgrat que el seu grau de penetració en la nostra societat, aquests dispositius presenten nous reptes i barreres a superar per a les persones discapacitades.

Si bé és cert que els telèfons intel·ligents moderns compten amb moltes opcions d'accessibilitat, aquestes capacitats dels dispositius no sempre són garantia suficient d'accessibilitat en les aplicacions. Cal que els desenvolupadors adaptin i integrin aquestes capacitats d'accessibilitat en les seves aplicacions.

Tenint en compte les característiques de portabilitat i el grau d'independència que els telèfons intel·ligents permeten, el desenvolupament d'interfícies basades en visió en dispositius mòbils és una alternativa molt interessant que poc contribuir a reduir aquesta problemàtica.

La recerca en el desenvolupament d'interfícies d'interacció natural en el cas de dispositius mòbils ha estat objecte d'estudi per multitud d'investigadors amb diferents aproximacions, que van des de l'ús de marcadors [14], la detecció de trets característics de la cara (*head-tracking*) [15-16], la detecció de la mirada (*eye-gazing tracking*) [17-18] i més recentment l'ús de xarxes neuronals profundes (*deep learning*) [19].

El desenvolupament d'interfícies basades en visió per computador en dispositius mòbils presenta una sèrie de reptes derivats de les característiques pròpies d'aquest tipus de dispositius.

Els principals reptes a afrontar són:

- Menor resolució d'imatge.
- Imatges de baixa qualitat a causa del soroll ambiental.
- La subjecció del dispositiu a la mà de l'usuari provoca problemes d'estabilització degut al moviment del dispositiu a la mà.
- Menor capacitat de càlcul en comparació amb entorns d'escriptori.

Tots aquests reptes exigeixen que el disseny d'algorismes de visió en dispositius mòbils s'hagi de realitzar amb màxima atenció a aspectes d'estabilitat, optimització de recursos i adaptabilitat.

En general el procés de construcció d'una interfície *head-tracking* es descompon en dos mòduls dependents en:

- Mòdul de detecció: En aquest mòdul es determina la localització de les àrees i trets d'interès de la imatge (en el cas d'aquest projecte cap, ulls i nas).
- Mòdul de seguiment: En aquest mòdul es localitza els objectes d'interès i es reporten els canvis en el flux de *frames* per dotar de continuïtat temporal a la interfície.

La integració de totes dues fases en un component d'interfície d'usuari facilita als desenvolupadors la integració de les interfícies basades en visió en les seves aplicacions seguint els patrons de disseny usats en el desenvolupament d'aplicacions per a dispositius mòbils.

Malgrat el gran interès del món acadèmic en l'exploració de noves formes d'interacció basades en visió per computador en plataformes mòbils, aquest interès no s'ha reflectit en el desenvolupament de solucions comercials, que permeti a les empreses de programari incorporar aquesta tecnologia en el desenvolupament de les seves aplicacions.

No obstant això, existeixen algunes solucions que han de ser esmentades com Eva Facial Mouse de la fundació Vodafone [26], TobiiPro de la startup Tobii [27], HeadMouse Nano de Origin Instruments [28] o EyeTribe [29] desenvolupament incorporat en 2017 al projecte Oculus de Facebook.

Eva Facial Mouse de la fundació Vodafone el sistema afegeix una nova funcionalitat de ratolí virtual al sistema operatiu. Aquesta aproximació presenta l'inconvenient que els desenvolupadors no perceben la tecnologia, i per tant no poden adaptar les seves aplicacions per adaptar-les a les persones discapacitades.

Tobbi Eye Tracker , EyeTribe o HeadMouse Nano, més orientades al món del videojoc, implementen aquesta funcionalitat mitjançant un dispositiu extern, el que limita la universalitat de la solució ja que prèviament cal que tots els usuaris de l'aplicació adquireixin el dispositiu.

L'objectiu ha aconseguir en aquest projecte és el desenvolupament d'una interfície *head-tracking*, que permeti a l'usuari interactuar amb les aplicacions sense necessitats afegits especials, i que pugui ser integrada fàcilment pels desenvolupadors en les seves aplicacions.

Analitzada l'estat actual de les solucions de programari d'interacció per visió per a dispositius mòbils, es considera que existeix un nínxol de mercat per al desenvolupament d'interfícies de visió basades en programari que puguin ser incorporades lliurement pels desenvolupadors en les seves aplicacions.

## 1.2 Objectius del Treball

L'objectiu d'aquest projecte és desenvolupar una interfície de visió per computador per a dispositius mòbils mitjançant tècniques de visió basades en el seguiment del cap o *head-tracking*.

El producte resultant deu ser capaç de reconèixer i interpretar la posició del cap de l'usuari i altres trets característics per a estimar la posició d'un element cursor en la pantalla del dispositiu mòbil.

Els objectius del projecte es poden resumir en els següents punts:

- Analitzar les tècniques de visió per computador per a detectar aquelles més adients per a la implementació de la interfície *head-tracking*.
- Dissenyar una interfície de *head-tracking* per a dispositius mòbils mitjançant la localització de diferents trets característics de la cara de la persona.
- Implementació d'un algorisme de *motion-tracking* sobre la base dels elements anteriorment detectats que permeti realitzar una estimació de la posició d'element punter en l'espai de la pantalla del dispositiu mòbil.
- Detecció de l'esdeveniment de *click* sobre la pantalla utilitzant detecció del gest de picada d'ullet en qualsevol dels ulls.
- Implementació d'un component d'interfície d'usuari (**GUI**) que permeti als desenvolupadors d'aplicacions aïllar-se dels aspectes concrets de sistema de visió.
- Desenvolupament d'una aplicació prova de concepte que permeti verificar la validesa de la solució.

El sistema es componen de diversos components funcionals interconnectats que implementen diverses parts i fases del procés de la interfície de seguiment.

Els mòduls tenen responsabilitats específiques en el procés de visió:

- **Aplicació prova de concepte:** aplicació nativa per a dispositius android que permet validar el funcionament d'algorisme de visió.
- **Component d'interfície de visió:** Component d'interfície d'usuari (*widget*) per a dispositius android que facilita la integració del sistema de *head-tracking* en les aplicacions. Les responsabilitats d'aquest component es troben:
  - Obtenció de la imatge de la càmera.
  - Comunicació amb el mòdul de visió per a l'enviament de la imatge obtinguda.
  - Mostrar una imatge de feedback a l'usuari amb informació afegida des del sistema de visió.
  - Recepció dels esdeveniments del sistema de visió.
- **Algorisme sistema de visió de *head-tracking*:** Implementació de l'algorisme de *head-tracking* per a dispositius mòbils android. Aquest algorisme es divideix en diferents mòduls:
  - Mòdul de tractament de la imatge: Implementa totes les operacions necessàries per a convertir, rectificar i normalitzar la imatge sense tractar obtinguda de la càmera.
  - Mòdul de detecció de cares: Implementa el classificador de detecció de cares.
  - Mòdul de detecció d'ulls: Implementa el classificador de detecció d'ulls.
  - Mòdul de detecció del nas: Implementa l'algorisme de detecció de nas.
  - Mòdul de *motion-tracking*: Implementa l'algorisme de detecció de moviment del nas com a element principal
- **Mòdul de detecció gest *wink*:** Implementa l'algorisme de detecció de gest picada d'ullet.
- **Módul del gestor d'events:** mòdul encarregat de generar els esdeveniments de detecció i posició cap a la capa de plataforma de dispositiu.

La Figura 1 mostra la divisió en mòduls de l'aplicació.

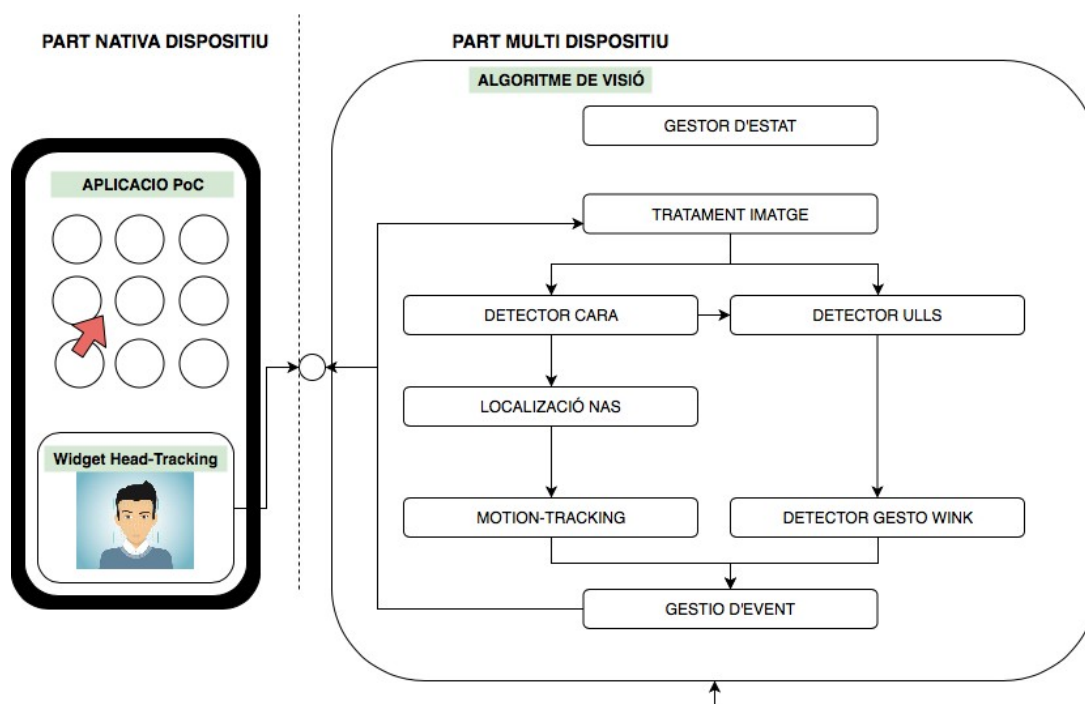


Figura 1: Components del sistema

### 1.3 Enfocament i mètode seguit

En una primera fase es procedirà a l'estudi de les tècniques de visió per computador d'interès per al projecte. En aquesta fase serà necessari familiaritzar-se amb les llibreries de visió OpenCV, així com el desenvolupament per a plataformes Android i la interfície de baix nivell JNI.

Una vegada adquirits el coneixements bàsics necessaris es dissenaran de cadascun dels mòduls del sistema, establint la seva funcionalitat i definint la comunicació entre ells.

Per portat a terme el disseny es procedirà a preparar els entorns de treball i les eines necessàries. En aquest cas serà necessàries eines per desenvolupament Android, com per exemple Android Studio, l'entorn de desenvolupament d'escriptori XCode.

En les primeres fases del desenvolupament l'esforç es focalitzarà en el desenvolupament de l'algorisme visió per *head-tracking* multiplataforma. En aquesta fase es procedirà a desenvolupar l'algorisme en l'entorn d'escriptori, per a posteriorment validar-lo en el dispositiu android. La validació de l'algorisme en l'entorn android generarà els canvis necessaris en el *widget* i en l'aplicació prototip.

Finalment, una vegada que l'algorisme es trobi estable i les funcionalitats bàsiques estiguin realitzades, l'esforç es pot focalitzar completar del widget i l'aplicació prototip. En aquesta fase s'implementarà els canvis necessaris per a la recollida de les dades del rendiment i les sessions dels usuari.

Per a garantir la consecució dels objectius es consideren els següents riscos i accions mitigadores o correctives exposats en la Taula 1.

*Taula 1: gestió de riscos del projecte*

RISC	IMPACTE	ACCIÓ
Retards en la temporització.	Alt	- Augmentar el nombre d'hores dedicades - Prioritzar les tasques - Eliminar funcionalitats no necessàries
Complexitat imprevista	Alt	- Buscar alternatives viables - Eliminar funcionalitats no necessàries
Problemes de rendiment algorisme	Mig	- Optimitzar el codi - Buscar implementacions alternatives - Eliminar funcionalitats no necessàries

#### 1.4 Planificació del Treball

Partint dels objectius estipulats en el Secció 1.2 el projecte s'han dividit en les següents fases (descrites en la Taula 2):

*Taula 2: planificació del treball*

FASE	TASQUES	ESTIMACIÓ
[TFG-1] Anàlisis	Anàlisis algorismes de visió	48 h.
	Anàlisis del classificadors face features	16 h.
	Anàlisis del algorismes motion-tracking	20 h.
	Redacció document anàlisis algorismes v1	4 h.
	Redacció document anàlisis algorismes v2	4h
	Redacció document anàlisis algorismes v2	4h
[TFG-2] Disseny	Disseny de components	48 h.
	Disseny interfície <i>head-tracking</i> .	20 h.
	Disseny interfície <i>head-tracking</i> (android).	8 h.
	Disseny interfície <i>head-tracking</i> (jni).	8 h.
	Disseny widget android.	4 h.
	Disseny prototip funcional.	16 h.
	Redacció document disseny components.	8 h.
[TFG-3] Implementació.	Implementació	224 h.
	Implementació prototip funcional.	40 h.

FASE	TASQUES	ESTIMACIÓ
	Test precisió.	16 h.
	Test robustesa.	16 h.
	Interfície usuari.	8 h.
	Interfície <i>head-tracking</i>	168 h.
	Daemon <i>head-tracking</i>	4 h.
	Inicialització camera + llibreries OpenCV	4 h.
	Bucle <i>head-tracking</i> (android)	16 h.
	Bucle <i>head-tracking</i> (JNI)	120 h.
	Tractament imatge raw camera	16 h.
	Classificador Face-Detection	8 h.
	Classificador Eye-Deteccion	16 h.
	Detecció posició nas	24 h.
	algorisme Motion-Tracking	24 h.
	Classificador Wink-Event	24 h.
	Gestor d'esdeveniments	16 h.
	Documentació algorisme <i>head-tracking</i>	8 h.
	Component android <i>head-tracking</i> (android)	16 h.
	Listener esdeveniments	8h
	Canvas Imatge de Feedback	8h.
[TFG-4] Validació.	Validació	16 h.
	Definició mètriques de validació	8h
	Validació subjecte #1-8	8h
[TFG-5] Tancament	Document conclusions	8 h.

## 1.5 Fites principals

Les fites principals del projecte son descrites en la Taula 3.

Taula 3: fites del projecte

DATA	DESCRIPCIÓ
27/03/2019	Document d'anàlisi de l'algorisme de visió per <i>head-tracking</i>
10/04/2019	Document de disseny – prototip funcional
29/04/2019	Llibreria OpenCV de l'algorisme
20/05/2019	Prototip funcional
02/05/2019	Memòria del projecte

## 1.6 Breu sumari de productes obtinguts

El projecte generarà els següents productes:

- **Aplicació prototip:** Aplicació android susceptible de ser instal·lada en dispositius compatibles amb les especificacions del projecte.
- **Component widget android:** Llibreria android que conté la implementació del component d'interfície d'usuari del sistema de *head-tracking*.

## 1.7 Estructura del document

El present document es troba dividit en 7 capítols, dividits a la vegada en diferents apartats.

- El primer capítol del document (capítol actual) on hem presentat una breu introducció del context d'aplicació, descrit el problema a resoldre, exposades les motivacions que m'han portat a la seva realització, definit el pla de treball i els diferents objectius a realitzar.
- El segon capítol del document presenta una breu descripció de la disciplina de visió per computador, posant especial èmfasi en les àrees concretes d'aplicació en aquest treball.
- El tercer capítol del document descriu el disseny de component del sistema, així com la interacció de les diferents parts per a la consecució dels objectius.
- El quart capítol del document descriu els detalls tècnics concrets de la implementació de l'algorisme de visió desenvolupat en aquest projecte.
- El cinquè capítol del document descriu el disseny de l'aplicació de prova de concepte implementada com a part de les fase de validació del projecte.
- El sisè capítol del document defineix les conclusions obtingudes després de la realització del projecte i proposa possibles treballs futurs.
- El setè capítol del document conté un glossari dels termes usats en la memòria.
- El vuitè capítol del document conté la bibliografia utilitzada en la memòria.



## 2. Interfícies de visió per computador

### 2.1. Visió per computador

La visió per computador (*computer vision* o **CV**) és el camp de la intel·ligència artificial que té com àrea d'estudi la capacitat d'imitar la visió natural per part dels ordinadors. L'objectiu és dotar a les màquines de la capacitat d'interpretar el contingut i les característiques d'una imatge.

És un camp d'estudi interdisciplinari que contempla el desenvolupament de la base científica i l'algorítmica per permetre els sistemes de visió siguin capaços d'aconseguir una compressió visual automatitzada.

La visió per computador com a disciplina de recerca té més de 50 anys en actiu. Va començar en els últims anys de la dècada de 1960 en universitats pioneres en el camp de la intel·ligència artificial, sorgint de la necessitat d'extreure els components tridimensionals continguts en una imatge amb la finalitat d'entendre'ls i construir coneixement.

Les dècades posteriors **van estar** marcades per múltiples avenços en el camp de la visió per computador. En el darrers temps, el desenvolupament de mètodes d'aprenentatge estadístic basats en característiques, l'aplicació de xarxes neuronals i l'aplicació de les xarxes neuronals profundes (*deep learning*) per a l'extracció de coneixement d'alt nivell, són algunes de les àrees en la qual s'han produït molts avenços.

Com a disciplina, la visió per computador està molt relacionada amb les disciplines del processament d'imatges, l'anàlisi d'imatges i la visió màquina. No obstant això els objectius de la visió per computador difereix dels d'aquestes disciplines i s'enfoca en:

- La detecció, segmentació, localització i reconeixement d'objectes en imatges.
- Determinar la presència d'objectes comuns presents en un grup d'imatges.
- El seguiment d'objectes en una seqüència d'imatges.
- La generació del model 3D de l'escena a partir del mapatge d'imatges.
- L'estimació de la postura de les persona en la imatge (*pose estimation*).
- La construcció de algorismes de cerca d'imatges per extracció de característiques.

Les aplicacions de la visió per computador en la indústria són múltiples, com podem veure en els exemples a continuació:

- En el sector de l'automoció s'usen sistemes de visió artificial industrial en les inspeccions de la fabricació i assemblatge de totes les peces dels vehicles.
- En el sector de l'alimentació s'usen sistemes de visió artificial industrial en els processos de control de qualitat dels productes alimentaris, com per exemple en el procés de selecció de fruites o el control de correcte tancament de les llaunes de conserva.
- Sistemes de visió industrial són usats en el control de qualitat en el cas de la indústria de l'embalatge.
- En el sector de l'electrònica s'usen activament sistemes de visió industrial en la manipulació i identificació dels diferents components, així com en el control de qualitat de soldadures i assemblatge de peces.
- Algorismes de visió per computador i reconeixement de patrons han estat usats àmpliament com a element d'identificació biomètrics.

## 2.2. Reconeixement de patrons

Una de les àrees de desenvolupament més rellevant dins el camp de la visió per computador és l'aplicació de tècniques de visió per al reconeixement de patrons. Aquesta disciplina sorgeix de la necessitat de compensar les limitacions de la ment humana en el reconeixement de patrons en entorns amb gran quantitat d'informació.

La determinació de la presència d'un objecte o un conjunt d'objectes amb característiques comunes requereix l'assignació d'etiquetes sobre la base de l'extracció d'atributs característics que actuen com a mètriques.

Un dels camps d'aplicació més comú dels algorismes de reconeixement de patrons és la detecció facial. La detecció facial ha estat àrea d'interès constant de l'entorn acadèmic i de les indústries, permetent així que, fruit d'aquest esforç combinat, en l'actualitat comptem amb sistemes de detecció facial robustos i eficaços.

La detecció facial és l'aplicació de patrons de reconeixements a una imatge a fi de determinar la presència del rostre, la seva posició, orientació i escala. Com en el cas d'aquest projecte, la combinació dels algorismes de detecció de la cara amb l'aplicació de tècniques de seguiment i estimació permetrà la sincronització dels moviments del cap de l'usuari amb un element 2D/3D situat en l'escena a representar.

Els sistemes de reconeixement facial presenten una sèrie de reptes i dificultats que fan que siguin difícils d'implementar en la pràctica i que els resultats obtinguts no sempre siguin els esperats. Les causes d'aquesta dificultat són diverses: l'angle del rostre respecte a l'eix de la càmera, les condicions d'il·luminació de l'escena, la presència d'ulleres, borrissol facial o longitud del cabell, són alguns exemples.

Diverses estratègies han estat utilitzades per a la implementació de sistemes de detecció facial:

- Detecció facial basada en trets.
- Detecció facial basada en models.
- Detecció facial basada en aparença.
- Detecció facial basada en aprenentatge automàtic.

La detecció facial basada en trets es basa en l'explotació de les propietats de la imatge com per exemple la presència del to de la pell o la detecció de la geometria general del rostre. L'ús del to de pell per a la implementació de la detecció de cares [25] permet la construcció de classificadors de forma senzilla i ràpida sobre la base del contingut cromàtic de la imatge.

Existeixen diversos espais de color per a la representació del color en les imatges, cadascun amb els seus avantatges i desavantatges.

El model RGB [30] representa el color mitjançant la combinació additiva de tres plans de color independents corresponent al color vermell, verd i blau. És un model de color poc intuïtiu, ja que no permet distingir fàcilment els matisos del color en la forma que els humans els percebem i el seu ús no sempre resulta adequat en sistemes de detecció. La Figura 2 mostra el model additiu RGB .

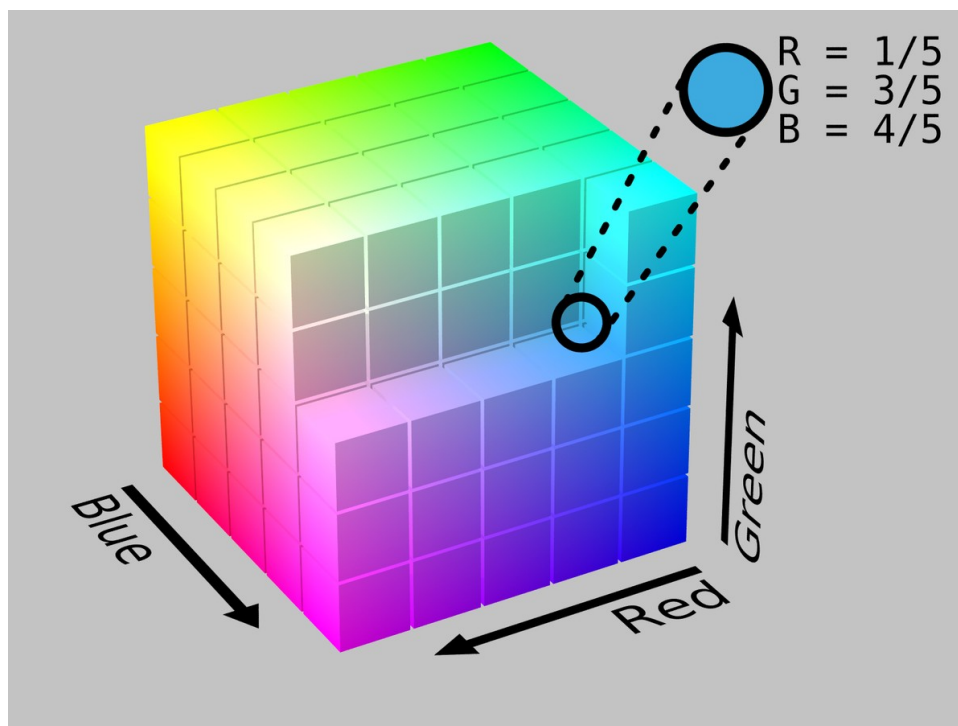


Figura 2: espai de color RGB

L'espai de color HSV [31] s'obté mitjançant una transformació no lineal de l'espai de color RGB i presenta la virtut d'aproximar-se millor a la visió humana. El pla H representa el to amb el color blanc en el seu centre, el pla S representa la saturació del to de color i el pla V defineix la lluentor o valor d'il·luminació. Aquest espai de color és molt usat sobretot per a avaluar les

condicions d'il·luminació de la imatge a causa de la separació de la component del valor d'il·luminació de la informació cromàtica de l'escena. La Figura 3 mostra el model HSV .

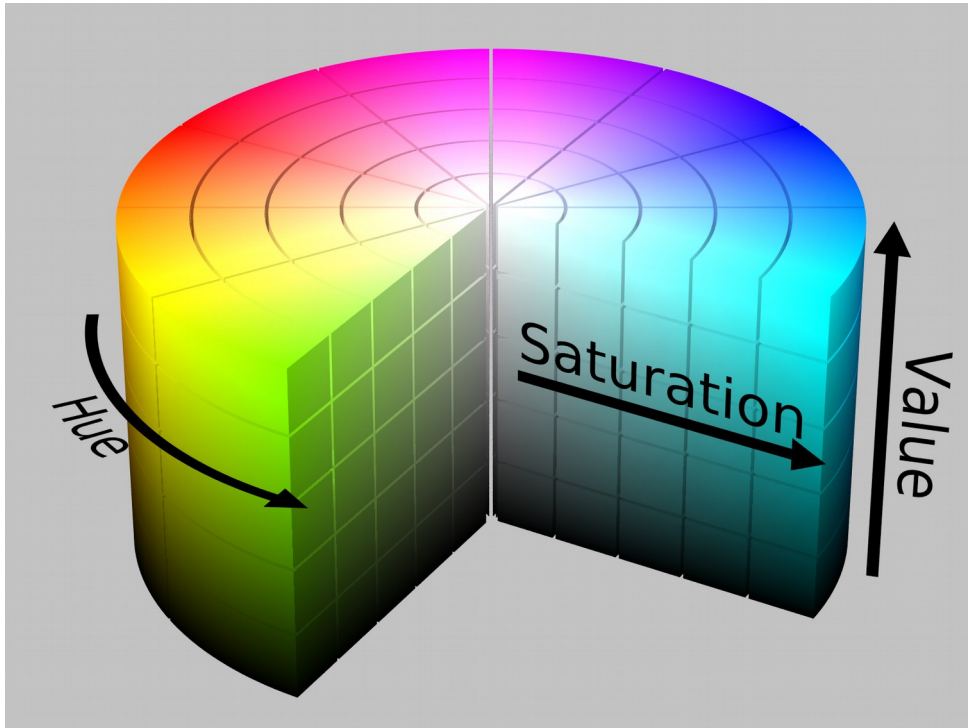


Figura 3: espai de color HSV

El espai de color YCbCr [32] elimina en gran part la redundància de l'espai de color RGB en la transmissió i difusió d'imatges de forma digital. El pla Y representa la component lumínica en una escala de grises, mentre que Cb i Cr representa el valor cromàtic corresponent al to blau i vermell respectivament. La Figura 4 mostra el model YCbCr.

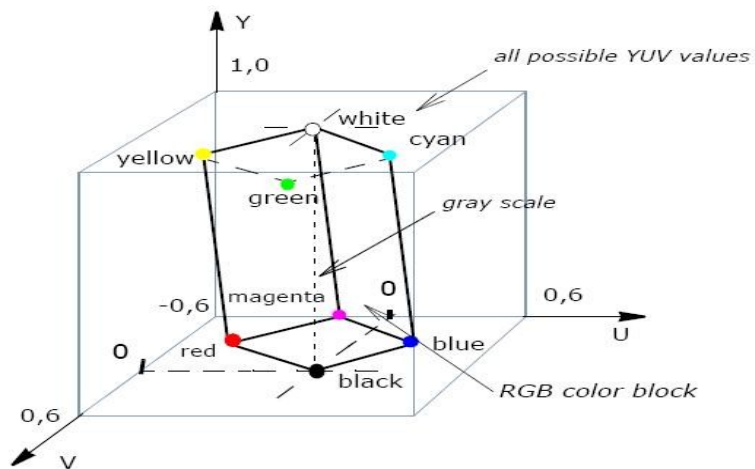


Figura 4: espai de color YcbCr

Tant els espais de color HSV com YCbCr han estat usats en la implementació de detectors facials basats en color amb una bona taxa d'encerts.

Els sistemes de detecció facial basada en models [24] es basen en l'obtenció d'informació geometria de la imatge, tals com vèrtexs, cantonades, línies i polígons. Aquesta informació geomètrica és comparada amb el model geomètric de referència representatiu del rostre.

En la pràctica aquest tipus de detector no ofereixen resultats òptims, pel fet que mínimes variacions en la imatge provoquen fragmentació dels contorns detectats. Aquesta fragmentació té com a conseqüència fallades en la detecció de vèrtexs i cantonades en zones claus o determinants de la imatge.

Els sistemes de detecció facial basats en aparença es fonamenten en les tècniques de cerca de coincidències (*template matching*). Les imatge de la cara són representades en un espai vectorial en la qual es realitza una cerca de concordances respecte als models de referència. Tècniques de detecció modernes com Fisherfaces o Eigenfaces són usades per a la implementació de detectors facials amb excel·lent resultats [33][34].

L'ús de tècniques d'aprenentatge automàtic (*machine learning*) com a xarxes neuronals han estat aplicades amb èxit en el desenvolupament de sistemes de detecció facial basats en aprenentatge automàtic.

Un dels algorismes d'aprenentatge automàtic més rellevants i popular és el mètode Viola & Jones. Aquest mètode va ser desenvolupat en 2001 per Michael Jones i Paul Viola [35][36].

Des de la seva publicació, aquest mètode va suposar un canvi molt important respecte a les tècniques de detecció de patrons usades a causa de la seva baixa taxa d'error i el seu excel·lent rendiment.

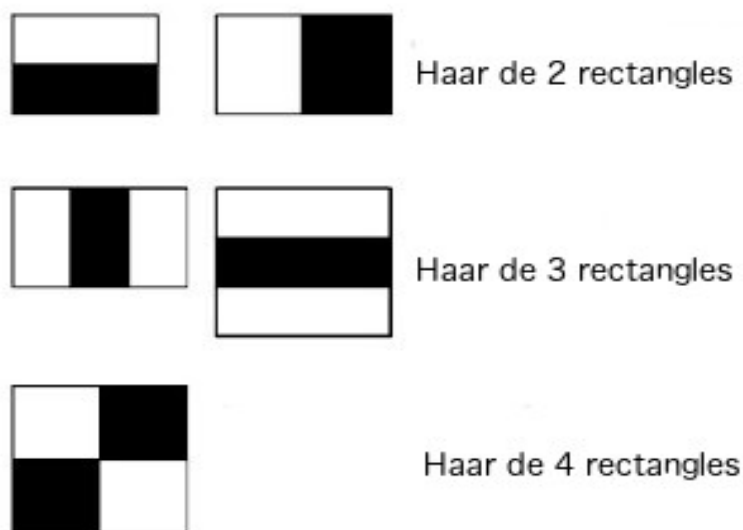
L'algorisme del mètode Viola & Jones presenta tres contribucions importants en l'àrea de la detecció de patrons.

- Representació integral de la imatge. Aquesta representació permet el còmput de les característiques de Haar de manera molt òptima.
- L'ús d'eficients classificadors febles de tipus *Adaboost* [37]. Aquests classificadors permeten seleccionar característiques concretes dins d'un conjunt enorme de característiques possibles.
- Combinació en cascada dels classificadors febles. Aquest esquema permet que la detecció sigui molt eficient en la tasca de descartar les característiques i zones que no contenen informació d'utilitat.

La construcció dels classificadors febles associats al mètode Viola & Jones es basen en l'extracció de característiques de Haar dels següents tipus:

- Característiques de Haar de 2 rectangles que avaluen la presència de vèrtexs horitzontals i verticals (Figura 5).

- Característiques de Haar de 3 rectangles que avaluen la presència de línies horitzontals i verticals (Figura 5).
- Característiques de Haar de 4 rectangles que avaluen l'avalua la presència de línies diagonals (Figura 5).



*Figura 5: característiques de Haar*

Les característiques rectangulars de Haar poden ser computades ràpidament gràcies a una representació intermèdia de la imatge anomenada imatge integral [20]. Una imatge integral és una estructura de dades que permet la suma ràpida de regions d'aquesta.

El mètode Viola & Jones utilitza classificadors febles AdaBoost dissenyats per a seleccionar una característica rectangular de Haar concreta. Per a cada característica es determina el valor llindar que garanteix una taxa de detecció pròxima al 99,9% al cost de tenir una taxa de rebuig o falsos positius pròxima al 50%.

Un detector de cares basat en el mètode de Viola & Jones usa aquests classificadors febles AdaBoost en el context d'una estructura d'arbre de classificació per descart en cascada on cada node és un classificador AdaBoost multiarbre.

Aquesta estructura en cascada es configura com un arbre degradat en la qual en cada nivell es pren una única decisió. L'estructura completa de classificadors febles així organitzada es comporta un classificador fort amb una taxa elevada d'incert. La Figura 6 mostra una representació de l'estructura d'arbre de classificació per descart en cascada de Viola& Jones.

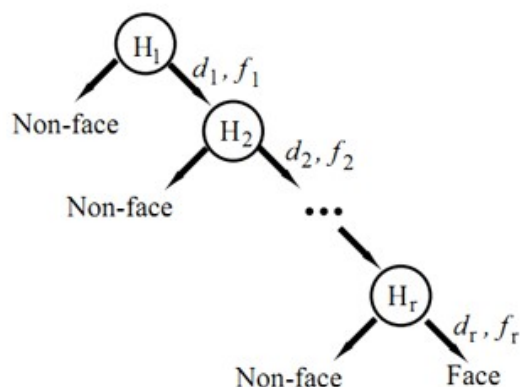


Figura 6: estructura en cascada

## 2.2. Interfícies *head-tracking*

La implementació d'un sistema de seguiment facial o *head-tracking* es pot descompondre de forma genèrica en l'execució concatenada de diversos passos.

La Figura 7 mostra la descomposició de les fases de la interfície *head-tracking*

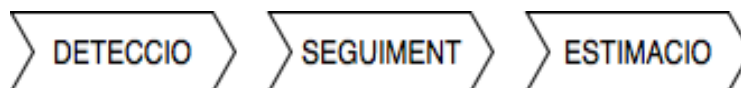


Figura 7: fases interfície *head-tracking* genèrica

En la fase de detecció s'aplica la detecció de cares per a determinar la presència, posició i escalat d'aquesta. En cas d'èxit del detector s'ha d'iniciar un procés de seguiment (*optical flow*) que registri els canvis de posició de la cara detectada. Sobre els canvis de posició registrats s'ha d'aplicar una estratègia de filtratge que anul·li els efectes de sorolls per a obtenir una successió de posicions estables.

Les tècniques de determinació del flux òptic es divideixen en dos grups:

- Tècniques de flux òptic dens, com els mètodes Meanshift [42] i Camshift [43].
- Tècniques de flux òptic dispers, com el mètode Lukas-Kanade [38].

El mètode de flux òptic dens es basa en l'estimació del moviment per a tots els punts d'una regió determinada. El resultat del mètode és un vector de desplaçament en cadascun dels punts que pot ser usat per a estimar el moviment. En la pràctica calcular el flux òptic no és tasca senzilla, per a cada punt el sistema s'han de trobar similituds a les regions veïnes en el següent *frame*.

Els algorismes de seguiment dens com Meanshift i Camshift presenten inconvenients que limiten la seva aplicació en el desenvolupament d'interfícies *head-tracking*. Aquests algorismes estan basats en la detecció de la tonalitat

de color. L'extracció del to exacte de la pell no és una tasca trivial i limita la detecció a persones amb color de pell que s'ajusti als rangs establerts.

Els algorismes de flux òptic dispers en canvi realitzen l'estimació del moviment mitjançant la cerca de punts característics de la imatge en els següents *frames*. Aquests algorismes són més eficients i robustos, al reduir el conjunt de punts de cerca en cada *frame*.

L'algorisme de flux òptic dispers més usat és el mètode de Lukas-Kanade [38]. Aquest mètode de cerca de flux de desplaçament per finestra de desplaçament utilitza trets característics de la imatge.

El mètode de Lukas-Kanade es fonamenta en 3 hipòtesis bàsiques:

- Hipòtesi de consistència de la lluentor del píxel en *frames* consecutius que exposa que en *frames* consecutius el valor de la intensitat de la llum en el punt es constant.
- Hipòtesi de persistència temporal que exposa que només es produeixen petits moviments entre *frames* consecutius.
- Hipòtesi de coherència espacial que exposa que els punts veïns d'una imatge presenten moviments similars.

El mètode Lukas-Kanade resol les equacions bàsiques de flux òptic per tots els píxels veïns, aplicant el criteri de quadrats mínims.

L'esquema de cerca Lukas-Kanade presenta el desavantatge de que només aquells desplaçaments que puguin ser fitats per les dimensions de la finestra de desplaçament són detectats. Conseqüència d'aquesta limitació es va proposar una versió millorada del mètode, aquesta versió es denomina mètode Lukas-Kanade piramidal [44].

La fase d'estimació del moviment és l'encarregada de filtrar les dades de seguiment obtingudes en la fase anterior. En un entorn ideal els resultats obtinguts en la fase de seguiment serien directament aplicables en la interfície d'usuari, en la pràctica això no és sempre possible. Les imperfeccions en el procés d'adquisició de la imatge introdueix soroll en el nostre procés de seguiment, aquest soroll provoca desviacions en el valor de seguiment.

L'element clau per a minimitzar l'efecte del soroll present en la fase de seguiment és la construcció d'un model d'estimació robust que tingui en compte els valors de seguiment obtinguts i els efectes del soroll en el sistema.

La tasca d'estimació del moviment es divideix en dues fases.

- Fase de predicció.
- Fase de correcció.



En la fase de predicció el model usa la informació d'estats anteriors per a determinar la següent posició possible. En la fase de correcció la posició obtinguda és incorporada al model per a ajustar els seus estats interns.

Un dels models d'estimació més usats és el filtre de Kalman [39][40]. La idea que fonamenta el filtre de Kalman és que sota unes premisses fortes però raonables és possible construir un model de l'estat actual del sistema sobre la base de la història passada que maximitza la probabilitat de predicció de posicions futures.

Tres premisses fortes però raonables són requerides per a la construcció del model del filtre de Kalman.

- El sistema a modelar és lineal.
- El soroll en l'obtenció de mesures és blanc.
- El soroll és gaussià per naturalesa.

El filtre de Kalman és un estimador que integra la informació de sobre l'estat d'un sistema, la informació sobre la dinàmica del sistema i la informació obtinguda per observació del sistema.

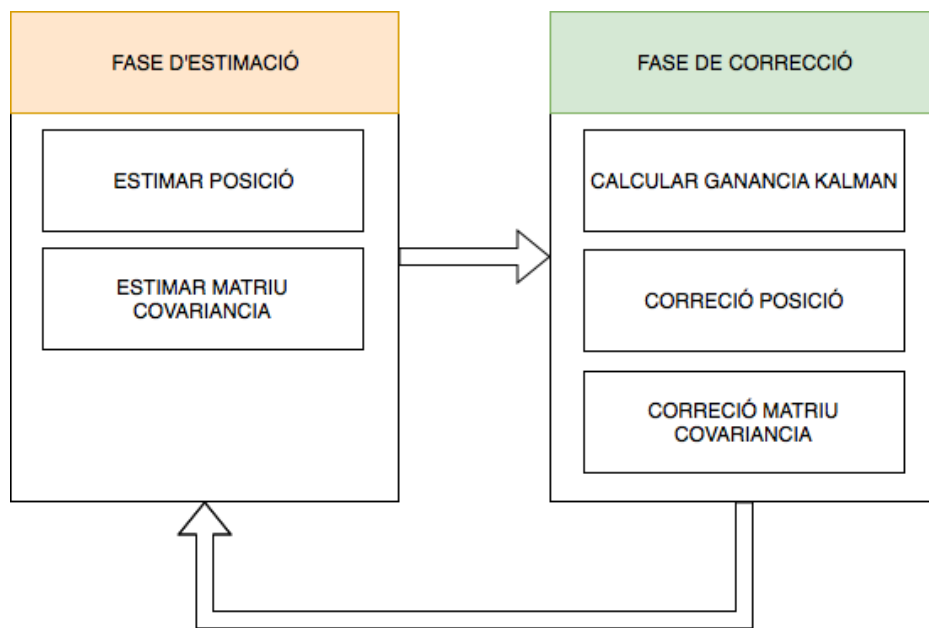
L'estat del sistema es representa com un vector que conté les variables que considerem rellevants, en el nostre cas la posició. Les variables rellevants són representades per una combinació de dos valors: mitjana gaussiana i covariància. La mitjana gaussiana representa l'estat de major probabilitat i la covariància que representa la incertesa sobre la mitjana.

Amb aquesta informació, en cada iteració el sistema estima la nova posició maximitzant la probabilitat de la nova posició donada el mesurament previ, tenint en compte l'estat previ i en grau d'incertesa de la variable a estimar.

En la pràctica el model del filtre de Kalman es compon de dues fases:

- Fase de predicció
- Fase de correcció.

En la fase de predicció es realitza després de l'obtenció d'un nou valor de mesurament però abans de l'incorporar-lo al sistema. En la fase de correcció s'apliquen les equacions del model del sistema per a obtenir una estimació corregida sobre la base de la informació d'estat anterior. La Figura 8 mostra la relació entre totes dues fases en el model de Kalman.



*Figura 8: relació de fases filtre de Kalman*

### 3. Disseny del sistema

En aquesta secció procedirem a analitzar i detallar els diversos components que engloben el sistema i com es relacionen entre ells per a la consecució de l'objectiu definit.

L'objectiu principal del projecte és el desenvolupament d'una interfície d'usuari per *head-tracking* per a dispositius mòbils. Aquest criteri de portabilitat determina una de les principals decisions de disseny aplicada: el disseny ha de permetre que el sistema sigui portable a noves plataformes. Per a aconseguir aquest objectiu és necessari minimitzar la capa dependent de les plataforma concreta.

El sistema es troba dividit en dos blocs o capes de disseny, cadascuna d'aquestes amb unes responsabilitats definides i diferenciades:

- capa de plataforma de la plataforma, encarregada dels detalls particulars de cada plataforma.
- Capa portable implementada sobre una base comuna utilitzant les llibreries OpenCV.

Aquest esquema presenta dos diversos avantatges importants:

- El sistema pot ser portat a altres plataformes per a les quals la capa comuna sigui compatible.
- L'esforç d'implementar el sistema en una nova plataforma es minimitza.

La figura 9 a continuació mostra els principals components del sistema i les seves relacions.

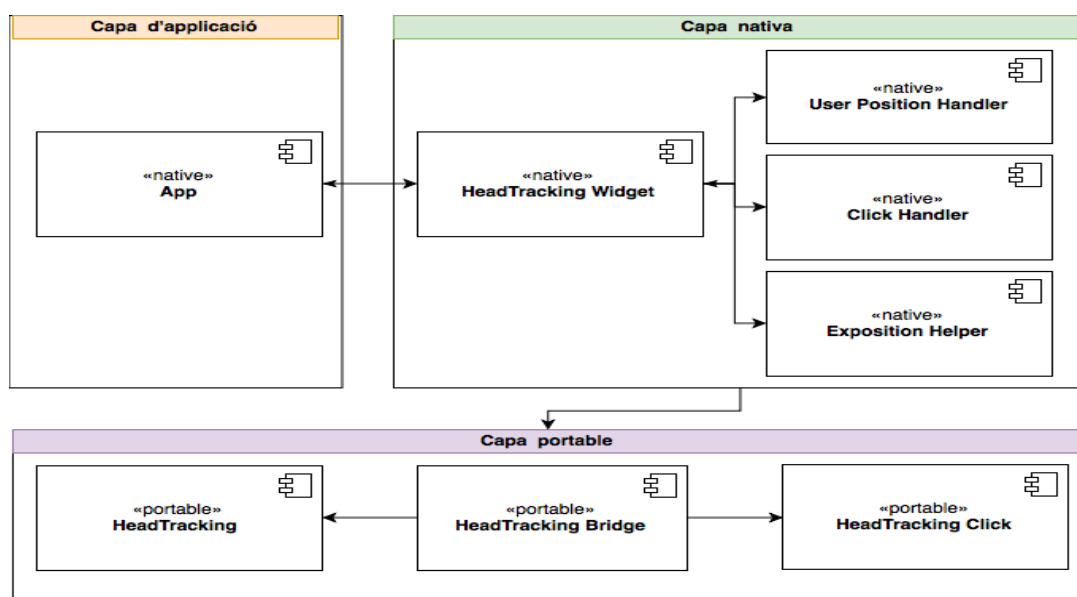


Figura 9: components del sistema

### 3.1. Component Widget.

Aquest component de la capa de plataforma encapsula les funcionalitats d'alt nivell de la interfície *head-tracking*.

Les responsabilitats d'aquest component són:

- Comunicació amb la capa portable de la interfície *head-tracking*.
- Gestió del resultat del processament de l'algorisme de *head-tracking* i generació dels esdeveniments de canvi d'estat, posició del punter de seguiment i generació de l'esdeveniment de *click*.
- Comunicació amb el component de càlcul de valor d'exposició de l'escena.

Aquest component, a través dels seus mètodes públics presenta la interfície que ha de ser implementada pels desenvolupadors d'aplicacions per comunicar-se amb el component que implementa l'algorisme *head-tracking*, els detalls de l'algorisme de *head-tracking* són objecte d'estudi en el Capítol 4.

És responsabilitat d'aquest component enviar el *frame* d'imatge obtingut des de la càmera del dispositiu al component de la capa de portabilitat que implementa l'algorisme de *head-tracking*.

La resposta del procés de tractament del *frame* per part del component de *head-tracking* de la capa de portabilitat és tractat en aquest component per a generar els esdeveniments corresponents que serveixin per a informar la capa d'aplicació.

La comunicació entre la capa d'aplicació i el component Giny es realitza mitjançant una comunicació per esdeveniments. L'esquema de comunicació per esdeveniments permet al component *Widget* informar la capa d'aplicació dels següents canvis en l'estat de l'algorisme de visió:

- Esdeveniment d'estat de sistema «en repòs» (idle).
- Esdeveniment d'estat de sistema «calibrant» (calibrating).
- Esdeveniment d'estat de sistema «en seguiment» (tracking).
- Esdeveniment de canvi de posició del punter virtual.
- Esdeveniment de gest de *click*.
- Esdeveniment de valoració de grau d'exposició.
- Esdeveniment de valoració de la qualitat del posicionament de l'usuari.

El component *Widget* s'ajuda de diversos components auxiliars per a implementar les funcionalitats necessàries de la capa de plataforma:

- Component *UserPositionHandler*: és el component encarregat de transformar i adaptar les coordenades de *frame* obtinguda per l'algorisme de seguiment a coordenades de pantalla per a ser usades en la capa d'aplicació.

- Component *ClickHandler*: és el component encarregat d'implementar l'esdeveniment de *click* per temps (*dwell-time*). Si el desenvolupador habilita aquest comportament, quant l'usuari manté la posició del cursor per un període de temps definit el sistema genera un esdeveniment de *click* sobre la posició mantinguda.
- Component *ExpositionHandler*: és el component encarregat d'interpretar la informació del valor d'exposició a partir del *frame* de la càmera i comunicar-lo a la capa d'aplicació.

### 3.2. Component HeadTrackingBridge

Aquest component actua com a nexa d'unió entre les capes nativa del dispositiu i la capa portable, responsable de la implementació de l'algorisme de *head-tracking*.

És responsabilitat d'aquest component realitzar les crides necessàries a la resta de components de la capa de portabilitat, així com mantenir tota la informació d'estat necessària per al correcte funcionament del sistema.

El component *HeadTrackingBridge* es comunica amb els següents elements de la capa de portabilitat:

- Component *HeadTracking*: És el component encarregat de determinar la posició del cursor a partir del *frame* obtingut de la càmera.
- Component *HeadTrackingClick*: És el component encarregat de determinar l'ocurrència del gestor de picada d'ullet (*wink*) a partir de la informació dels ulls estreta del *frame* de la càmera.
- Component *Exposition*: És el component encarregat de calcular el valor d'exposició de l'escena a partir del *frame* obtingut de la càmera.

### 3.3 Component HeadTracking

Aquest component implementa l'algorisme de seguiment *head-tracking*, els detalls d'aquest algorisme són objecte d'estudi en el Capítol 4.

A partir del *frame* obtingut de la càmera s'invoca el mètode de processament del *frame* d'aquest component per a determinar la posició del punter imaginari i actualitzar la informació de l'estat intern del component.

Tota la lògica associada a l'algorisme de visió *head-tracking* es troba continguda en aquest component sense necessitat de dependències externes, més enllà de les dependències necessàries de les funcions i estructures de dades de la llibreria OpenCV.

### 3.4. Component HeadTrackingClick

És el component encarregat de processar el *frame* obtingut de la càmera de determinar l'existència de la gest de picada d'ullet per part de l'usuari. Els

detalls de la implementació d'aquest algorisme són objecte d'estudi en el Capítol 4.

Tota la lògica associada a l'algorisme es troba continguda en aquest component sense necessitat de dependències externes, més enllà de les dependències necessàries de les funcions i estructures de dades de la llibreria OpenCV.

### **3.4. Component Exposition**

Aquest component és l'encarregat d'estimar el valor d'exposició de l'escena sobre la base del *frame* obtingut de la càmera mitjançant una escala pròpia. El valor obtingut té caràcter informatiu amb l'objectiu d'informar de l'existència de condicions d'il·luminació no propícies per al bon funcionament de l'algorisme de seguiment.

La Figura 10 mostra el diagrama de les classes esmentades i les seves relacions, segons s'ha descrit en el capítol.

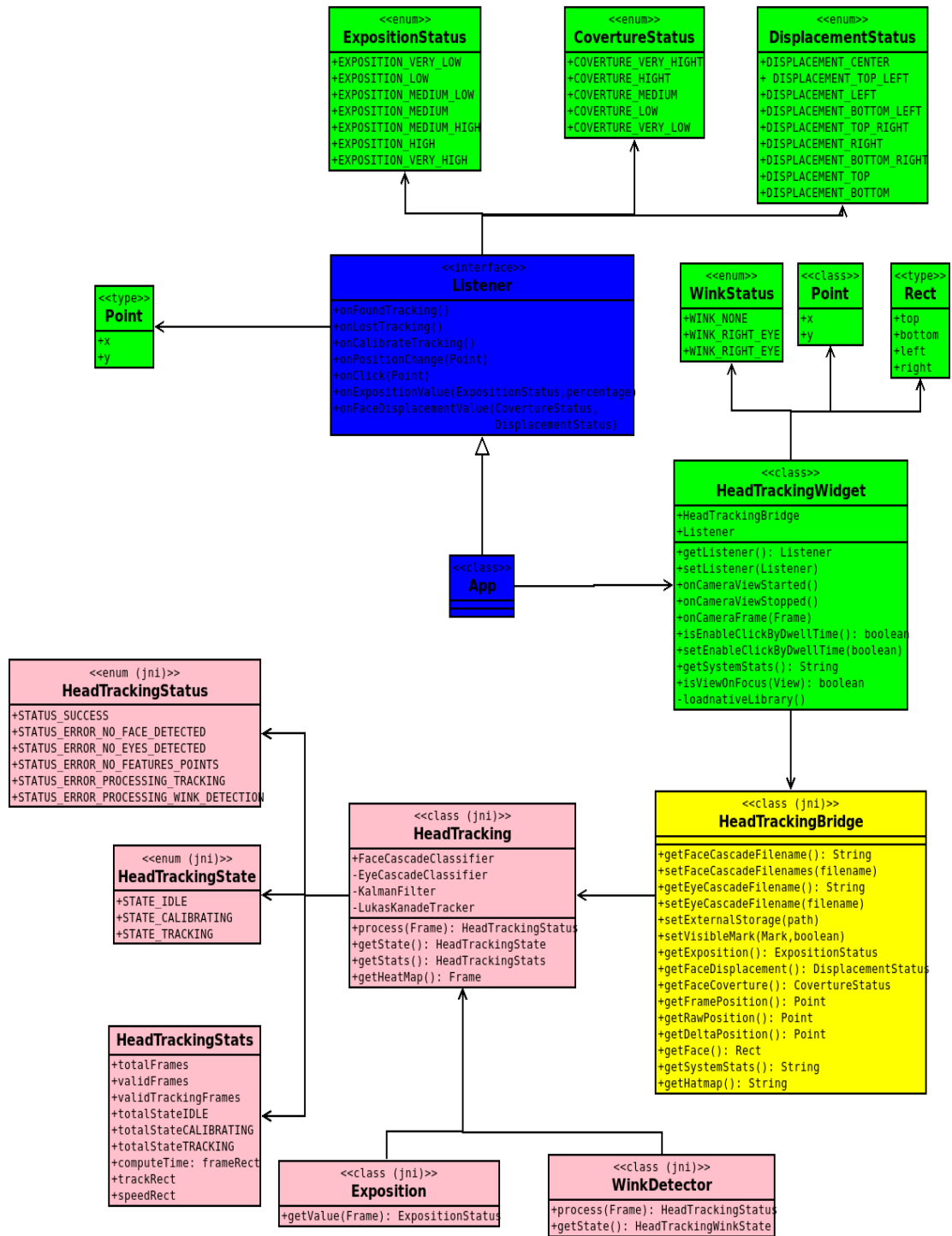


Figura 10: diagrama de classes

La figura 11 mostra el diagrama de seqüència i el flux de cridades entre els components, segons s'ha descrit en el capítol.

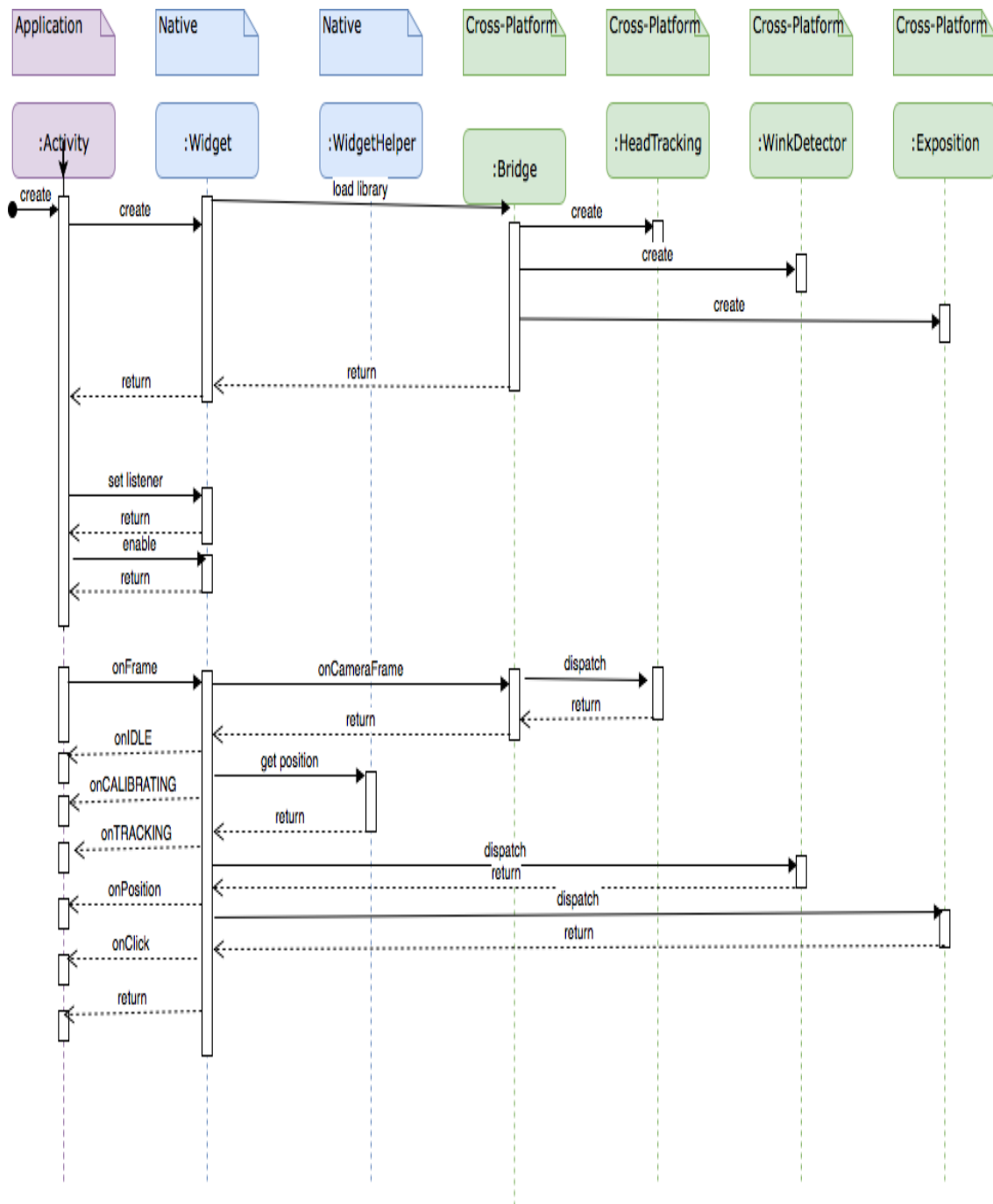


Figura 11: diagrama de seqüència



## 4. Algorisme de visió

En aquesta secció s'exposaran les diverses fases que componen l'algorisme de visió de la interfície *head-tracking* implementat en aquest projecte. La lògica de l'algorisme de visió es troba implementada íntegrament en la capa de portabilitat.

L'objecte d'aquest algorisme és determinar la posició d'un punt imaginari en la zona de màxima intensitat del nas que es correspon amb la part punxeguda coneguda com a àpex a partir de les imatges obtingudes des de la càmera del dispositiu.

El seguiment dels canvis de posició d'aquest punt són transmesos a la capa superior de la interfície d'usuari encarregada de transformar aquest punt en coordenades de pantalla i d'implementar a través d'aquesta característica un punter virtual, objectiu final d'aquest projecte.

La Figura 12 mostra una representació d'alt nivell de l'algorisme.

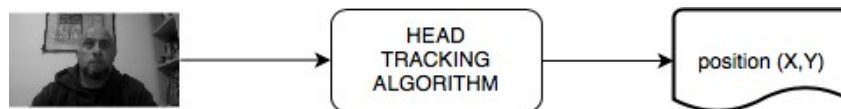


Figura 12: entrada i sortida de l'algorisme

L'algorisme de *head-tracking* accepta com a paràmetre d'entrada una imatge color o en escala de grisos i, presenta com a valor de sortida una estructura de dades amb la posició de punter imaginari del nas inferit.

Tal com s'especifica en la Secció 3.3 d'aquest document, l'algorisme de *head-tracking* es situa en la part més interna del sistema i s'ha implementant en C++ fent ús de la llibreria android JNI i de la llibreria de visió OpenCV en la versió 4.0.1.

L'algorisme es divideix en les següents fases seqüencials:

- Fase de detecció.
- Fase de seguiment.
- Fase d'estimació.
- Fase de transferència.

#### 4.1. Fase de detecció.

L'objectiu d'aquesta fase és determinar la presència dels punts característics del rostre (*features points*) adequats per a ser transferits a la següent fase, la fase de seguiment.

La Figura 13 resumeix els principals passos implementats en la fase de detecció.



Figura 13: fase de detecció

El primer pas de l'algorisme consisteix a adequar la imatge d'entrada al format esperat pel classificador, en el cas d'aquest projecte a escala de grisos. En cas que les imatges d'entrada estiguin en color, serà necessari transformar-les a escala de grisos.

Tot seguir es porta a terme la detecció de la cara, aquesta fase es realitza a mitjançant un classificador en cascada que utilitza el model de detecció frontal basat en característiques de Haar.

És important destacar en aquest punt que els models en cascada utilitzats són models de detecció frontal, això vol dir que han estat entrenats amb mostres de cares només en orientació frontal i que per tant no serà possible la detecció de cara quan l'usuari perd aquesta orientació. La figura 14 mostra diferents orientacions de la cara de l'usuari durant el procés de detecció.

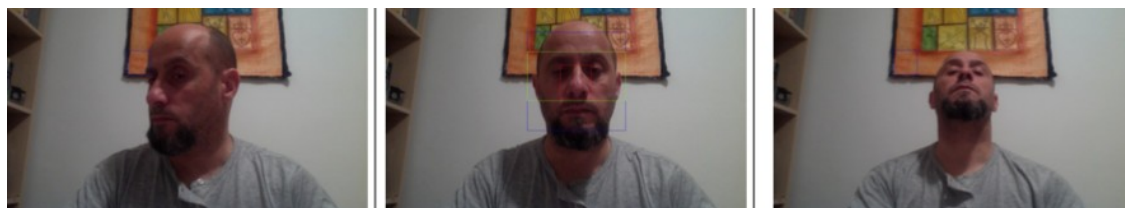


Figura 14: fallades de detecció frontal

En les proves realitzades s'observa que el classificador deixava de ser funcional si l'angle de rotació de l'usuari respecte a l'eix de la càmera supera els 20° sobre l'eix vertical (rotació *yaw*) o sobre l'eix horitzontal (rotació *pitch*).

En el cas concret d'aquest projecte es va considerar que aquests valors límit són acceptables a causa de les dimensions reduïdes de la pantalla en els dispositius mòbils, aquestes dimensions obliguen l'usuari ha col·locar-se en una posició pròxima i frontal al dispositiu per a no perdre la visió de la pantalla.

Els valors inicials de configuració del detector de cares van ser establerts mitjançant el mètode de prova-error amb validació de la mètrica de rendiment positiu (rendiment detector = frames amb cares detectades/total frames).

La Taula 4 mostra els valors per defecte dels paràmetres del classificador.

Taula 4: opcions de configuració del detector

DESCRIPCIÓ	PARÀMETRE	VALOR
Mida cara mínima	DETECTOR_FACE_MIN_OBJ_SIZE	10
Mida cara màxima	DETECTOR_FACE_MAX_OBJ_SIZE	400
Factor d'escalat piramidal	DETECTOR_FACE_SCALE_FACTOR	1.1
Valor coincidència nivell piràmide	DETECTOR_FACE_MIN_NEIGHBOURS	1
Valor màxim seguiment detector	DETECTOR_FACE_MAX_TRACK_LIFETIME	20
Valor mínim <i>frames</i> detecció	DETECTOR_FACE_MIN_DETECTION_PERIOD	1

Per la presència de diverses persones o a causa de falsos positius és possible que el detector retorni diverses cares, en aquest cas cal realitzar un procés de filtrat per a seleccionar la més adient aplicant els següents criteris:

- Criteri de major dimensió.
- Criteri de posicionament en relació al centre del *frame*.

La Figura 15 mostra la detecció de diverses cares en la imatge i els errors per fals positiu.

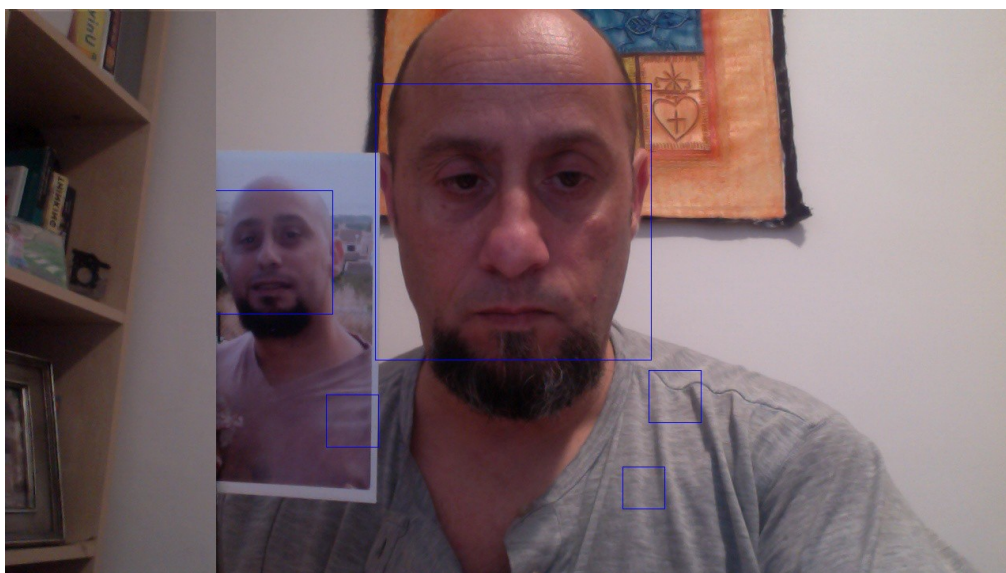


Figura 15: error de detecció múltiple

Amb la informació de la regió de la cara obtinguda, s'aplica regles antropomòrfiques [41] i es determina una regió fitada que contingui el nas. Aquesta regió fitada per al nas està definida per la regla de  $\frac{1}{3}$  del àrea de la cara detectada en l'eix vertical i horitzontal.

La figura 16 mostra una representació de les regles antropomòrfiques del rostre [41].

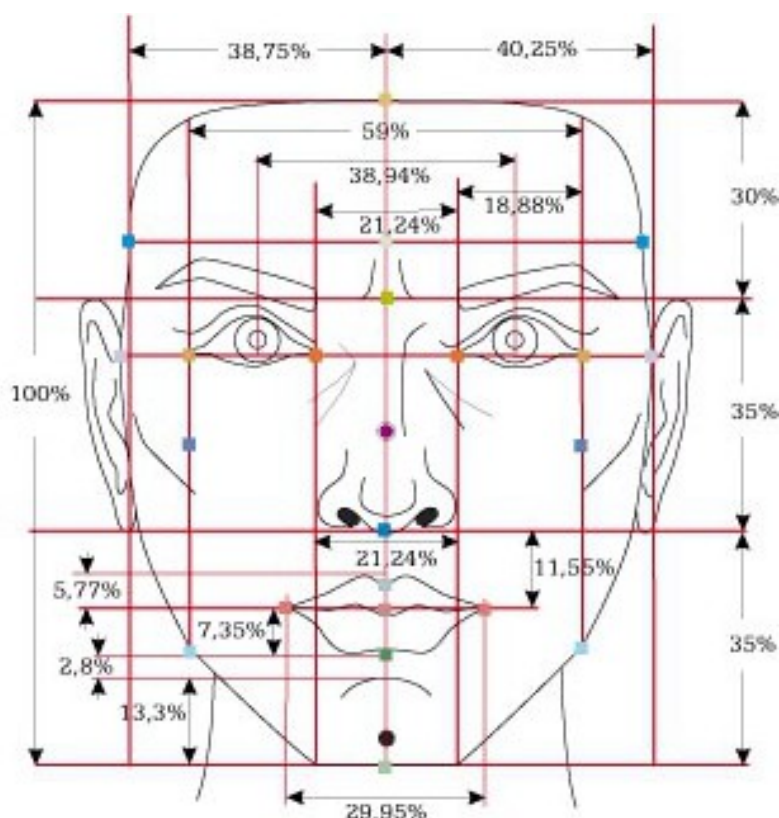


Figura 16: regles antropomòrfiques del rostre

En aquest punt s'aplica un segon classificador per a detectar la posició dels ulls. La informació obtinguda d'aquest classificador s'utilitza per ajustar l'àrea del nas amb l'objectiu d'obtenir una àrea més petita i precisa. La Figura 17 mostra el l'àrea de nas abans i després de l'ajust de l'àrea del nas.

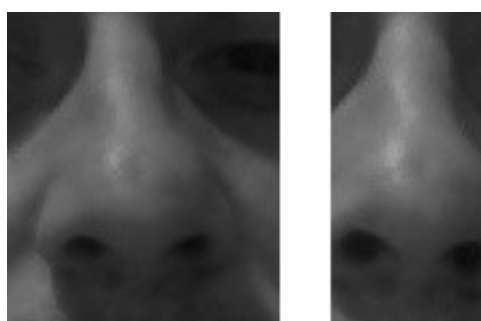


Figura 17: regió del nas afitada

L'àrea del nas obtinguda actua com a màscara d'acotació en el procés d'extracció de punts característics (*feature point*). La determinació d'aquests punts es realitza mitjançant les funcions OpenCV *goodFeaturesToTrack*.

Aquestes funcions retornen punts on es produeixen canvis bruscos en el gradient d'intensitat en la imatge. És per aquesta raó que el nas i en concret la punta o àpex, és un element susceptible d'acumular gran quantitat de punts d'alta intensitat.

Si aproximem el model d'il·luminació de l'escena per un model de reflectància de la superfície de Lambert [22] observem que la punta del nas pot ser modelada o aproximada per una mitja esfera. Aquesta mitja esfera presenta una superfície on es produeixen màxims d'intensitat en les zones on la direcció de la llum i direcció de la normal al punt de la superfície formen un angle de 180 graus.

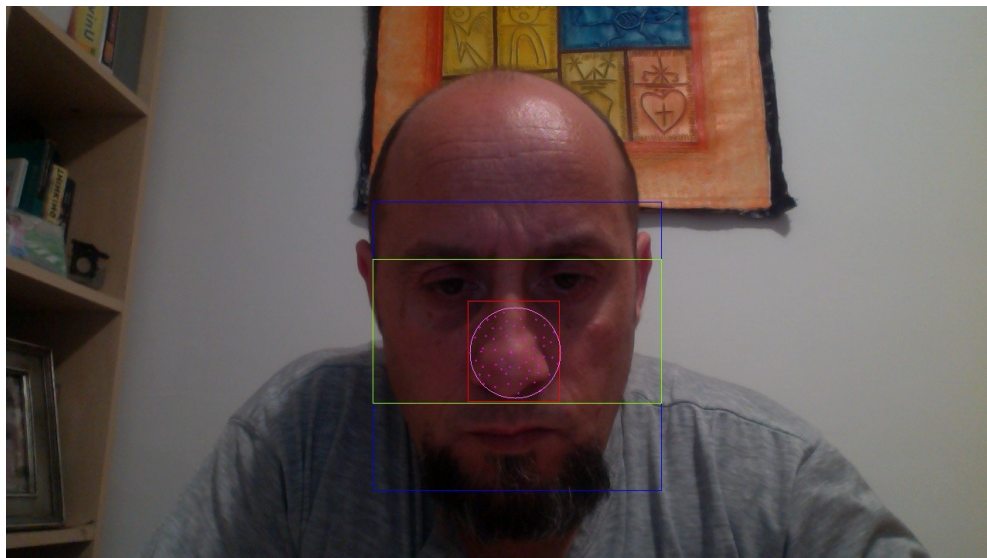
Els valors inicials de configuració del procediment d'extracció de punts característics es van establir pel mètode de prova-error amb validació del nombre de punts característics trobats. La Taula 5 mostra els valors per defecte dels paràmetres dels mètodes d'extracció de punts característics.

*Taula 5: opcions de configuració del càlcul de punts característics*

DESCRIPCIÓ	PARÀMETRE	VALOR
Màxim de punts característics	FEATURES_MAX_FEATURES	200
Numero de nivells de cerca	FEATURES_MAX_LEVEL	3
Mida d'un bloc mitjà	FEATURES_BLOCK_SIZE	3
Indica ús detector de Harris	FEATURES_HARRIS_DETECTOR	false
Paràmetre lliure del detector Harris	FEATURES_K	0.04
Qualitat mínima	FEATURES_QUALITY	0.01
Mínima distància euclidiana	FEATURES_MIN_DISTANCE	10
Iteracions màximes	FEATURES_CRITERIA_ITERATIONS	20
Precisió desitjada	FEATURES_CRITERIA_EPS	0.03

Resultat del procés anterior obtenim un llistat de punts característics (*feature points*) continguts a la regió del nas determinada. Estem interessant només en aquells que es troben en el centre de l'àrea del nas seleccionada, per aquesta raó s'aplica un procés de filtrat per distància al centre de la zona per a descartar aquells punts més allunyats.

La Figura 18 mostra el resultat del filtrat de punts per la distància al punt central.



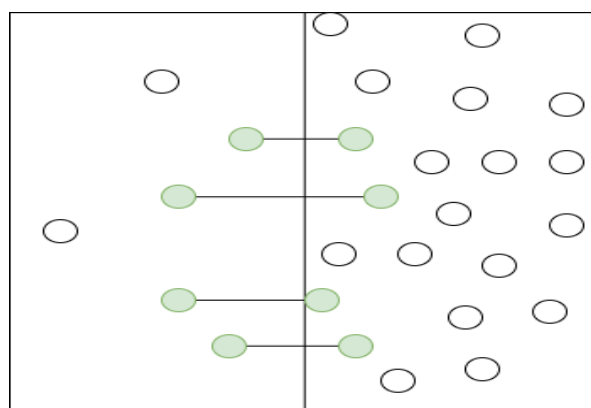
*Figura 18: filtratge de punts per distància del punt central*

A causa de les condicions d'il·luminació particulars de l'escena és freqüent que la direcció de la llum respecte a la punta del nas provoqui una major concentració de punts característics en una zona concreta del nas.

Amb l'objectiu d'evitar aquest biaix que provocaria un desplaçament de punt imaginari, que desitgem que se situï en la zona central de la punta del nas, realitzem un procés d'aparellament de punts per a ponderar la seva influència.

S'estableix un eix vertical sobre el punt mitjà de la regió del nas que divideix els punts oposats en dues meitats. S'aparellen els punts més pròxims de totes dues meitats i s'afegeixen a la col·lecció de punts candidats, al final del procés tots els punts sense aparellar són descartats.

La Figura 19 mostra l'esquema d'aparellament de punts.



*Figura 19: aparellament de punts*

## 4.2. Fase de seguiment.

L'objectiu d'aquesta fase és correlar la informació dels punts característics (*features points*) obtinguts en la fase de detecció amb la informació dels punts característica obtinguts en la iteració anterior per tal de determinar el desplaçament d'aquests punts. L'algorisme de seguiment utilitzat en aquesta fase és l'algorisme de Lukas-Kanade.

El resultat de l'aplicació de l'algorisme de flux òptic Lukas-Kanade ens retorna una llista de punts candidats, resultat del desplaçament dels punts característics en els *frames* consecutius.

Els valors inicials de configuració del procediment d'extracció de punts característics es van establir pel mètode de prova-error amb validació del nombre de punts característics trobats. La Taula 6 mostra els valors per defecte dels paràmetres del mètode de flux òptic Lukas-Kanade.

Taula 6: opcions de configuració de l'algorisme de Lukas-Kanade

DESCRIPCIÓ	PARÀMETRE	VALOR
Màxim de punts característics	LUKAS_KANADE_MAX_FEATURES	200
Numero de nivells de cerca	LUKAS_KANADE_MAX_LEVEL	3
Mida d'un bloc mitjà	LUKAS_KANADE_BLOCK_SIZE	3
Indica ús detector de Harris	LUKAS_KANADE_HARRIS_DETECTOR	false
Paràmetre lliure del detector Harris	LUKAS_KANADE_K	0.04
Qualitat mínima	LUKAS_KANADE_QUALITY	0.01
Mínima distància euclidiana	LUKAS_KANADE_MIN_DISTANCE	10
Finestra de cerca	LUKAS_KANADE_WINDOWS_SIZE	10
Regió morta al mig de la zona de cerca	LUKAS_KANADE_ZERO_ZONE	-1
Iteracions màximes	LUKAS_KANADE_CRITERIA_ITERATIONS	20
Precisió desitjada	LUKAS_KANADE_CRITERIA_EPS	0.03

La implementació utilitzada d'aquest algorisme no restringeix el procés a una zona fitada si no que es considera tot l'àrea del *frame*. Aquesta limitació provoca que es retornin punts que es poden trobar fora de la zona desitjada, fins i tot fora de la zona de la cara detectada. Per a eliminar aquesta limitació es descarten tots els punts que són fora de la regió  $\frac{1}{3}$  de la cara.

La Figura 20 mostra la detecció de punts característics en el procés de flux òptic que es troben fora de l'àrea del nas.



*Figura 20: detecció de punts fora de l'àrea fitada*

Amb el conjunt de punts característics fitats a la regió desitjada, la determinació de la posició corresponent al punt imaginari de seguiment es realitza mitjançant el càlcul de la mitjana aritmètica de la posició de cadascun dels punts característics obtinguts. El punt mitjà calculat és transferit a la fase de següent d'estimació.

#### **4.3. Fase d'estimació.**

El procés de determinació del punt mitjà candidat des de les fases anteriors no és un procés perfecte. És freqüent que es produeixin fortes variacions i comportament no desitjat fruit de condicions de l'entorn provocades per canvis d'il·luminació, influència de les ombres, oclusió temporal dels rostres, imperfeccions en els sensors o detecció de falsos positius en els classificadors.

Aquestes circumstàncies provoquen que la localització del punt mitjà presenti una certa inestabilitat. Per a solucionar aquest problema i dotar d'estabilitat al sistema es modela la trajectòria seguida pel punt mitjà, per a aconseguir-ho s'empra un element estimador de tipus filtre de Kalman.



La figura 21 mostra en blanc la posició del punt mitjà candidat i la posició obtingut després d'aplicar el filtre de Kalman.

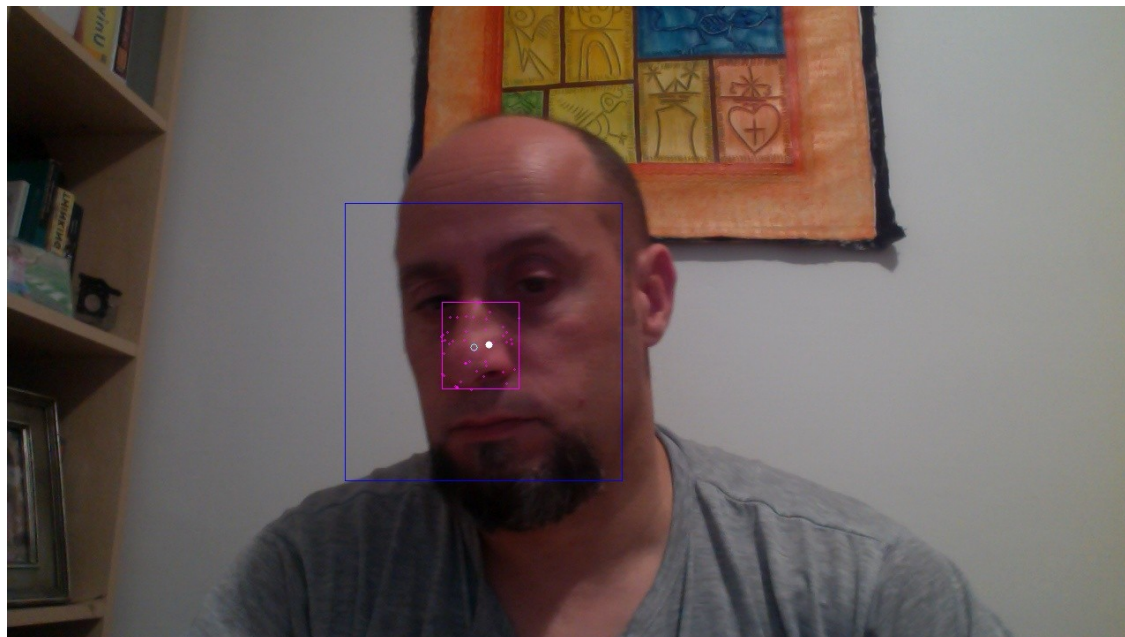


Figura 21: punt mitjà i punt mitjà de Kalman

Els valors inicials de configuració del procediment d'estabilització del punt mig es van establir pel mètode de prova-error mitjançant observació del rendiment. La Taula 7 mostra els valors per defecte dels paràmetres del filtre de Kalman.

Taula 7: opcions de configuració del model del filtre de Kalman

DESCRIPCIÓ	ELEMENT	VALOR
Dimensionalitat del vector estat.	KALMAN_DYNAMIC_PARAMS	4
Dimensionalitat de la mesura.	KALMAN_MEASURE_PARAMS	2
Dimensionalitat del vector de control.	KALMAN_CONTROL_PARAMS	0
Valor covariància del soroll del procés	KALMAN_PARAM_PROCESS_NOISE	1000.0
Valor covariància del soroll del mesura	KALMAN_PARAM_MEASURE_NOISE	10000.0

#### 4.4. Fase de transferència.

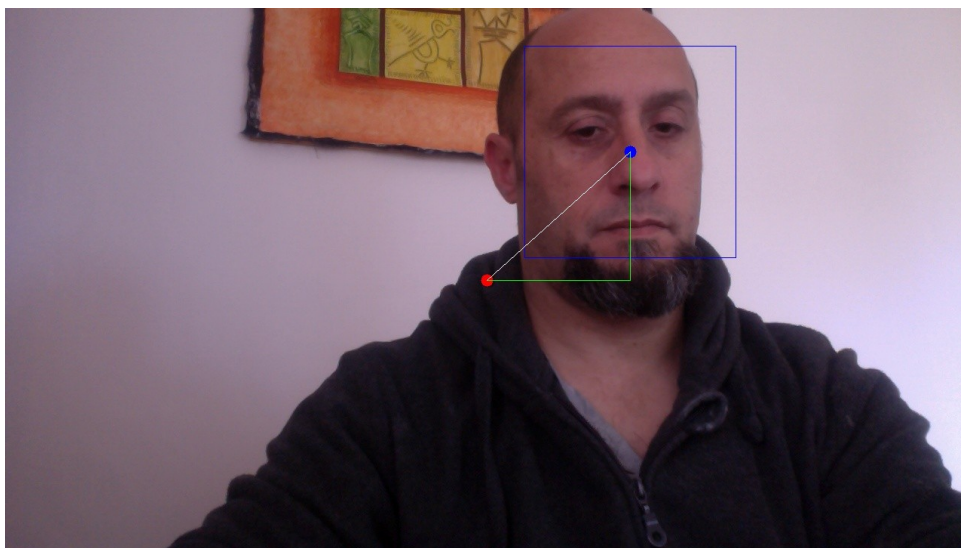
El resultat de les fases anteriors obtenim la posició estabilitzada del punt imaginari centrat en el centre de la punta del nas. Aquest punt és relatiu a les coordenades del *frame* d'entrada i per tant no és directament utilitzable des de l'aplicació.

Per a poder utilitzar com a punter en la interfície d'usuari el punt imaginari, és necessari trobar la correspondència d'aquesta posició en coordenades de pantalla.

Aquesta correspondència s'ha de portar a terme tenint en compte els següents factors:

- **Factor 1:** La posició de l'usuari en el *frame* obtingut de la càmera del dispositiu.
- **Factor 2:** Els desplaçaments de la posició de la càmera per moviments de la mà que sosté el dispositiu.
- **Factor 3:** La capacitat i rang articular de l'usuari.

Durant el procés de seguiment l'usuari no està sempre posicionat en el centre de la imatge, és freqüent que canviï la seva posició. Aquest desplaçament provoca que el centre de la cara es trobi desplaçat respecte al centre del *frame* (**factor 1**). La Figura 22 mostra el desplaçament del punt central de la cara respecte al centre del *frame*.



*Figura 22: posició de l'usuari desplaçada del centre*

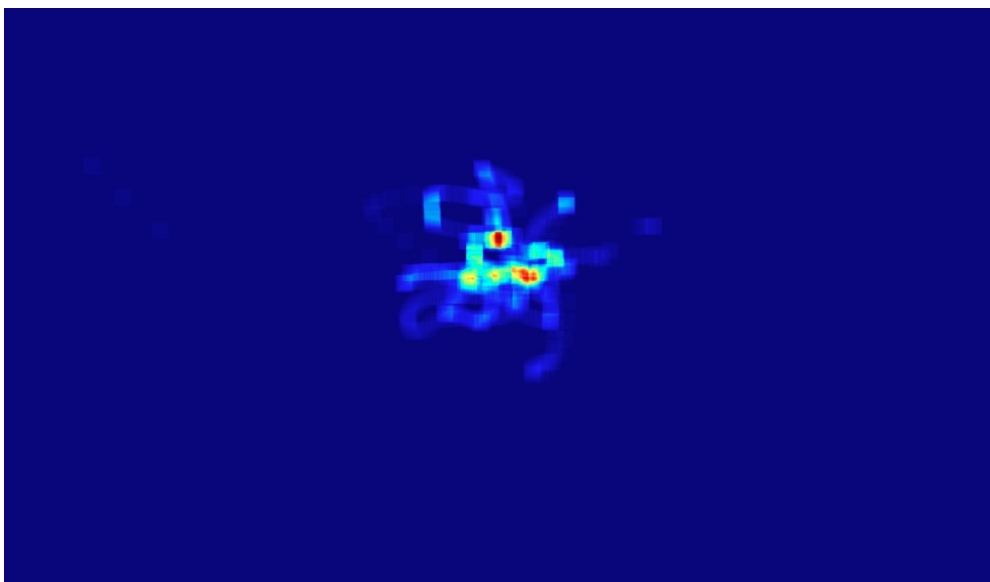
Durant el procés de seguiment, l'usuari subjecta el dispositiu mòbil a la seva mà, els canvis de posició de la mà produeixen canvis de la posició de la càmera respecte a l'usuari (**factor 2**). Aquest canvi de posició relativa de la càmera es tradueix en desplaçaments del punt central de la cara en relació al centre del *frame*. La Figura 23 mostra l'efecte del canvi de posició relatiu de la cambra i el desplaçament del punt central que provoca.



*Figura 23: desplaçament de la càmera*

Les limitacions del moviment de l'usuari juntament amb les limitacions de rotació del classificador frontal usat en el projecte, provoquen que tan sols una petita part del *frame* de la càmera pugui ser utilitzada per a representar el moviment del punter imaginari situat en el punt anteriorment obtingut.

La figura 24 mostra el mapa de calor del moviment de punt de seguiment imaginari en el centre del nas durant un interval de temps.



*Figura 24: mapa de calor moviment del nas*

La funció de transferència del punt de seguiment, haurà de tenir en compte els factors 1,2,3 per establir les coordenades de forma adient. Això realitza mitjançant la següent fórmula:

$$\begin{aligned} \text{screen}_x &= \text{scale}_x * \text{punt}_x * \text{screen}_{\text{width}} \\ \text{screen}_y &= \text{scale}_y * \text{punt}_y * \text{screen}_{\text{height}} \end{aligned}$$

El valor del factor d'escalat no pot ser configurat de forma fixa sinó que ha de ser ajustat de forma dinàmica per a adaptar-se a les necessitats de l'usuari.

Per a això s'estableix el següent comportament:

- El factor d'escalat en cadascun dels eixos i direccions s'estableix de forma dinàmica en funció del desplaçament mínim i màxim de l'usuari.
- La determinació del factor d'escalat es realitza mitjançant una relació lineal dels desplaçaments mínims i màxims en el rang de valors [6-10].
- Aquest esquema permet que el factor d'escalat sigui mínim en cas de petits desplaçaments, facilitant a l'usuari el procés de selecció d'elements en la pantalla. Per contra si l'usuari realitza desplaçaments grans el factor d'escalat és màxim permetent a l'usuari desplaçar-se ràpidament per la pantalla.
- En cas de detectar canvis de desplaçament del centre de la cara o canvis en la dimensió de la cara detectada que superin un valor llindar, el sistema estableix el factor d'escalat al valor inicial, forçant d'aquesta

forma que es torni a realitzar el procediment d'ajust descrit en el punt anterior.

La Figura 25 mostra la relació dels espais de coordenades de l'algorisme, del *frame* i de la pantalla.

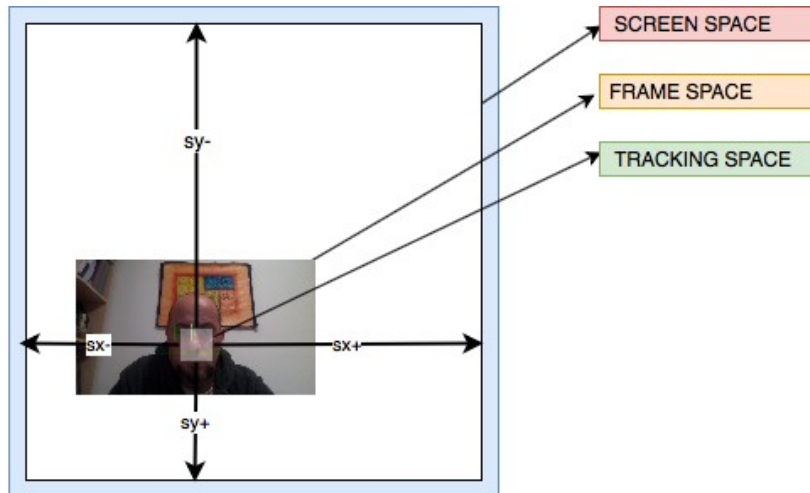


Figura 25: relació espais de coordenades

#### 4.5. Estabilitat del procés del seguiment.

Com s'ha descrit en els punts anteriors, existeixen múltiples aspectes en les diferents fases de l'algorisme que poden ser causa potencial de la fallada en el procés de seguiments de l'algorisme. La influència d'aquests factors d'error condueixen al fet que el *frame* obtingut des de la càmera no sigui vàlid per al procés de seguiment.

Amb l'objectiu d'afegir robustesa a l'algorisme es defineix una finestra de temps, mesurada en número de *frames*, dins la qual l'anterior previsió del punt de seguiment sigui vàlida. Aquest esquema permet que el sistema sigui més estable RESPECTO a la pèrdua esporàdica de *frames*, permetent així fer front als errors provocats per aquest.

La figura 26 mostra una representació del funcionament de la finestra de temps en cas d'errors de processament del *frame*.

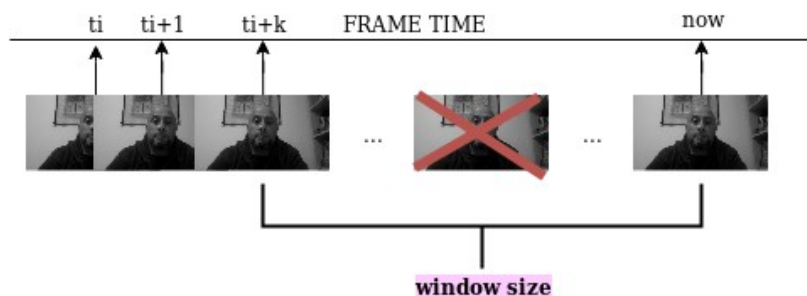


Figura 26: esquema de buffer de frames

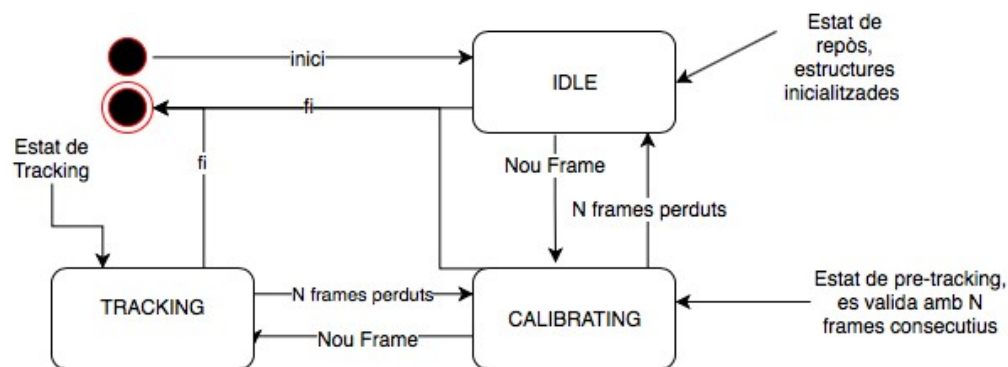
Perquè aquest esquema sigui útil cal que l'algorisme presenti un rendiment adequat de tal manera que la pèrdua de *frames* esporàdics no tingui impacte rellevant en el sistema.

A fi d'optimitzar el rendiment de l'algorisme s'ha establert diversos estats del sistema i distribuïdes les tasques necessàries a realitzar en cadascun dels estats.

Els estats definits en el sistema són:

- Estat de repòs (*idle*).
- Estat de calibratge (*calibrating*).
- Estat de seguiment (*tracking*).

La figura 27 mostra l'esquema de la màquina d'estats implementada en l'algorisme de visió.



En el estat de repòs (*idle*) no comptem amb la informació històrica necessària per a iniciar el procés de seguiment. L'arribada d'un *frame* vàlid segons les condicions anteriorment descrites en la fase de detecció provoca que s'iniciï el procés de determinació de l'àrea de nas i extracció de punts característics. Una vegada finalitzat aquest procés el sistema transita cap a l'estat de calibratge (*calibrating*).

L'objectiu d'estat de calibratge (*calibrating*) és obtenir informació que permeti iniciar el procés de seguiment descrit en la fase de seguiment i garantir que model d'estimació implementat pel filtre de Kalman convergeixi cap a un punt d'estabilitat.

Aquest estat es mantingut durant un numero determinat de *frames* igual a 10. Com s'ha esmentat anteriorment el sistema implementa una finestra de *frames* acumulats de tal manera que cada *frame* vàlid augmenti el valor d'aquesta finestra i cada *frame* no vàlid aquest valor decreix. En cas d'aconseguir el valor màxim de *frames*, en aquest cas igual a 10, el sistema transita cap a l'estat de seguiment (*tracking*), si el numero de *frames* arriba arriba a 0 el sistema torna a l'estat de repòs (*idle*).

La Figura 28 mostra un resum dels estats i les tasques que es duen a terme en cadascun d'ells.

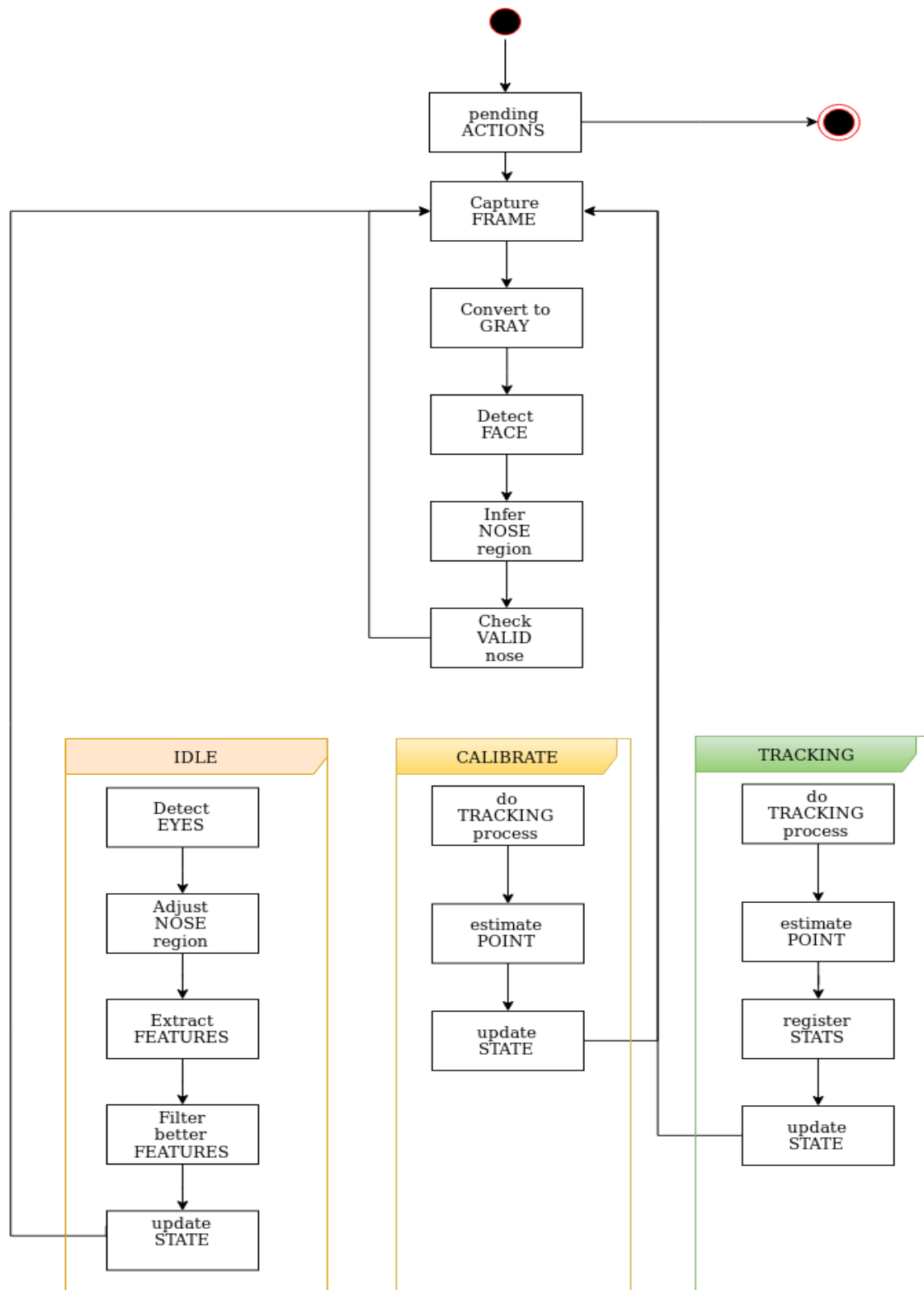


Figura 28: divisió de tasques per estat

La Figura 29 mostra un resum del flux dins de l'algorisme de *head-tracking* explicat en aquesta secció.

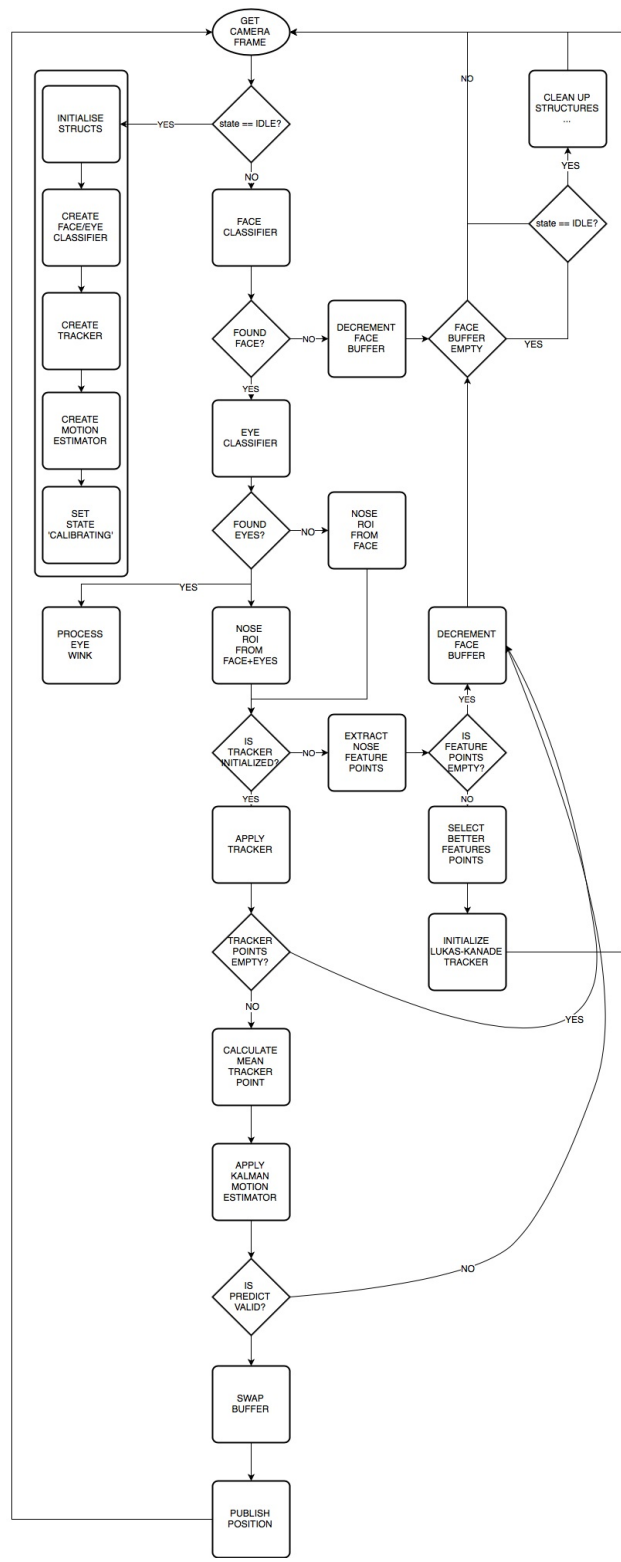


Figura 29: visió general de processos de l'algorisme

#### 4.6 Detecció de l'esdeveniment de *click*

Perquè el sistema implementi una interfície d'usuari útil cal que proveeixi de funcionalitat de selecció que permetin a l'usuari interactuar amb els elements de la interfície d'usuari.

En el cas dels dispositius mòbils aquesta interacció es produeix mitjançant les capacitats tàctils. De manera simple el sistema de seguiment *head-tracking* proveeix de funcionalitats similars de selecció de components.

Es faciliten dues implementacions d'aquesta funcionalitat, permetent així, al desenvolupador, triar aquella que millor s'ajusti a les necessitats de la seva aplicació.

Les dues opcions disponibles són:

- Interfície basada en la detecció del gest de picada d'ullet mitjançant tècniques de visió per computador (*wink detection*).
- Interfície basada en el temps de manteniment de la posició del punter de seguiment (*dwell-time*).

La detecció de la presència o absència del gest de picada d'ullet es basa en l'aproximació de la forma general de l'ull aplicant les següents premisses:

- Si l'usuari està tancant l'ull, la forma detectada d'aquest es pot aproximar mitjançant una línia.
- Si l'usuari té l'ull obert, la forma detectada d'aquest es pot aproximar mitjançant un polígon de 3 o més costats.

La Figura 30 mostra la divisió en tasques del procés de detecció del gest de picada d'ullet.



Figura 30: fases del algorisme detecció gesto *wink*

La Figura 31 mostra el flux en els diferents processos de l'algorisme de detecció del gest picada d'ullet.

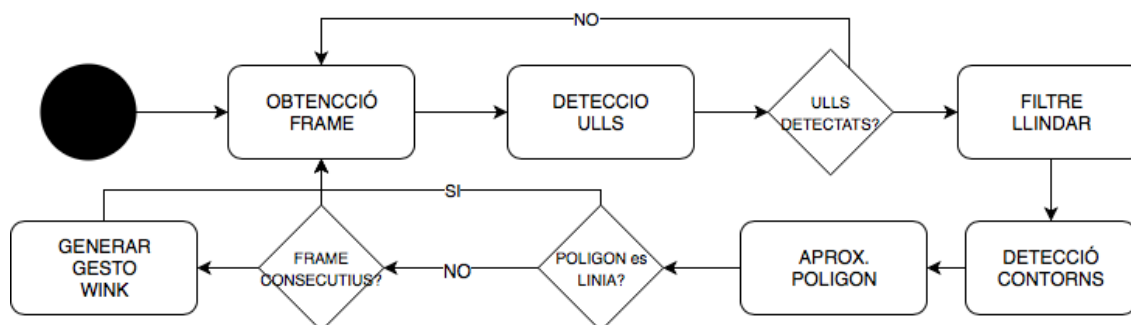


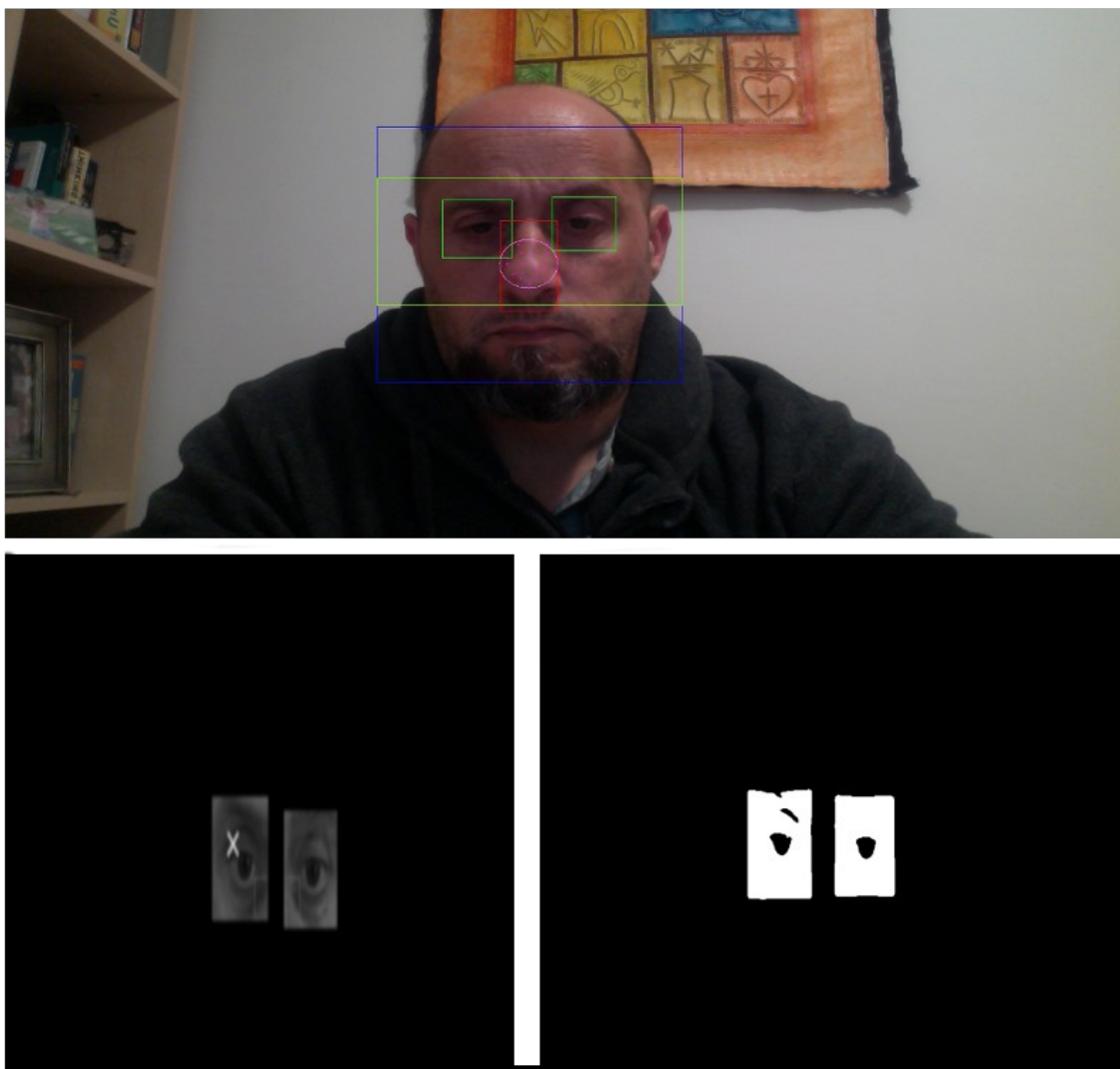
Figura 31: processos de l'algorisme detecció gesto *wink*



En la fase de detecció s'aplica un detector en cascada dels ulls per a obtenir les zones del *frame* on es troben situats. La posició de l'ulls detectats en relació a la dimensions de la cara detectada ens determina si es correspon amb l'ull esquerre o dret.

Sobre les zones detectades s'aplica un filtre de suavitzat de tipus gaussià per a minimitzar el soroll altament present en aquesta zona a causa de la seva morfologia. A continuació s'aplica un altre filtre, en aquest cas de tipus llindar binari per a aïllar les parts de l'ull de la resta de pell detectada pel classificador.

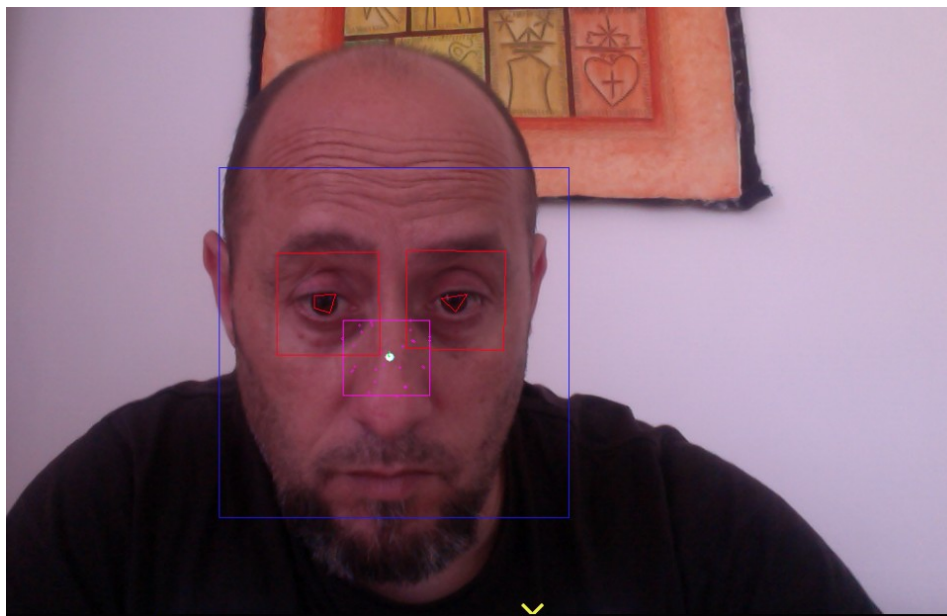
La Figura 32 mostra l'efecte de les tasques d'extracció de la zona dels ulls i filtrat per a obtenir una imatge binari d'aquests.



*Figura 32: tasques d'obtenció imatge binari dels ulls*

El resultat del filtratge anterior és una imatge binària en la qual es troba aïllada la geometria aproximada de l'ull. Sobre aquesta imatge es realitza els procediments d'extracció de contorns i d'aproximació poligonal.

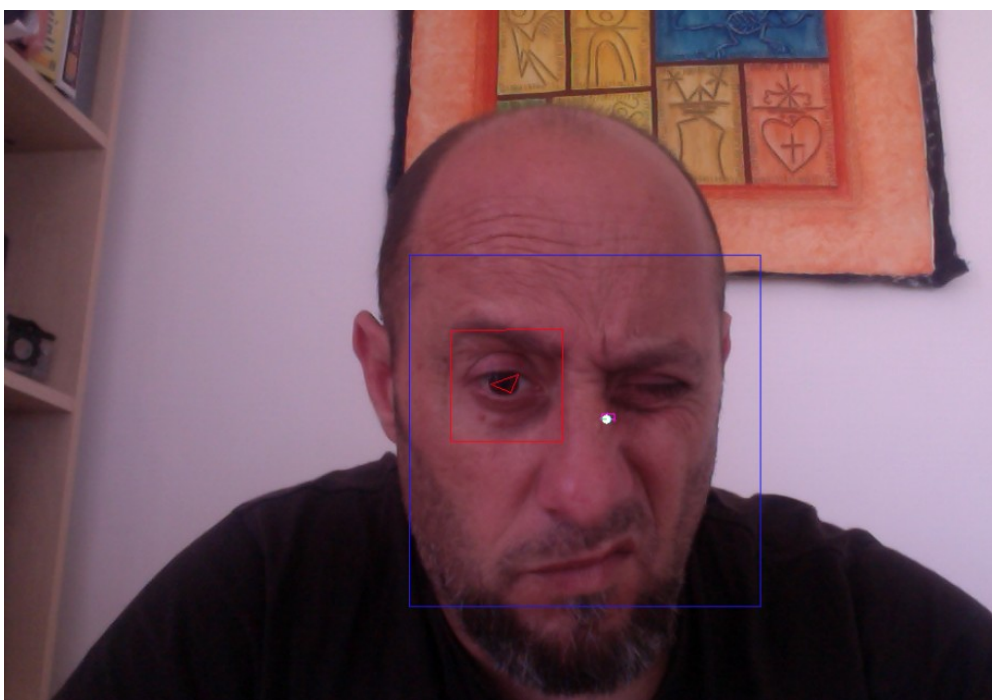
La Figura 33 mostra els polígons d'aproximació obtinguts.



*Figura 33: polígons d'aproximació als ulls*

Com a pas final de l'algorisme es verifica el nombre de vèrtexs del polígon obtingut per a determina si pot ser aproximat per una línia si l'usuari es troba tancant l'ull, o per un polígon de 3 o més vèrtexs si l'usuari té l'ull obert. En cas de detectar que l'usuari es troba tancant l'ull durant un període de temps determinat en *frames*, es genera un esdeveniment de tipus *click*.

La figura 34 mostra el resultat del procediment de detecció del gest de picada d'ullet.



*Figura 34: detecció del gest mitjançant polígons d'aproximació*

En les proves realitzades s'ha pogut comprovar que l'aplicació del classificador en cascada per a la detecció dels ulls presenta una elevada taxa d'error i de falsos negatius. Per aquesta raó es determina la necessitat d'una aproximació alternativa pel temps de permanència (*dwell-time*).

La figura 35 mostra un cas d'error per fals negatiu en la detecció dels ulls.

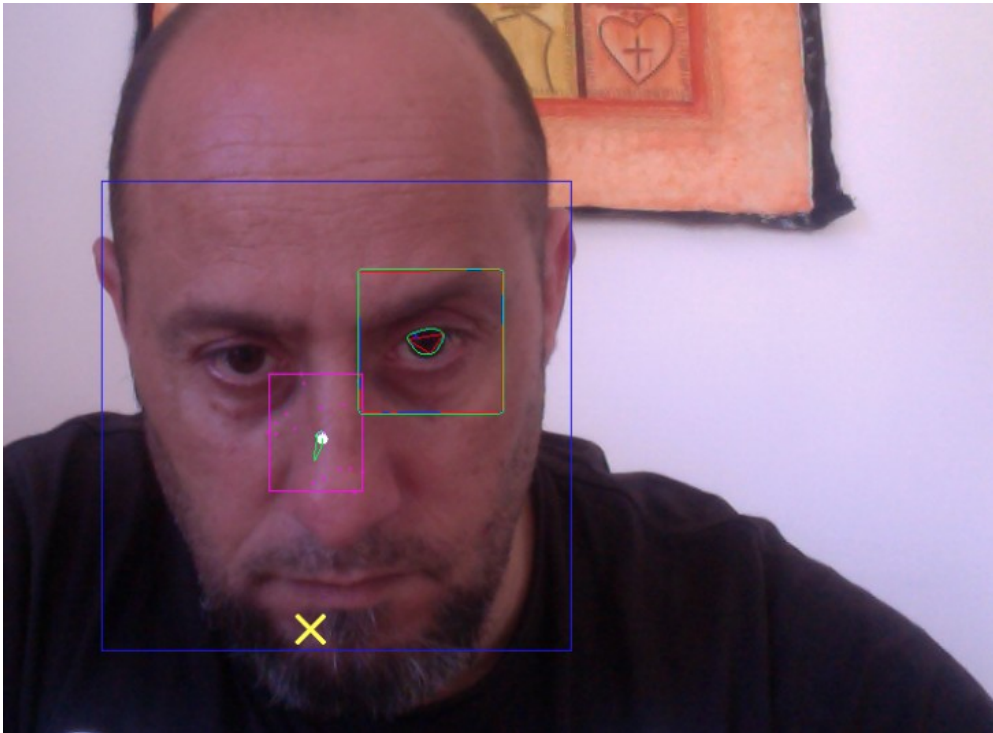


Figura 35: fals negatiu del detector d'ulls

La interfície *dwell-time* detecta el temps de manteniment de la posició actual del punter de seguiment per part de l'usuari, una vegada superat el valor llindar el sistema genera un esdeveniment de *click*. La Figura 36 mostra el flux del procediment de detecció de *click* per *dwell-time*.

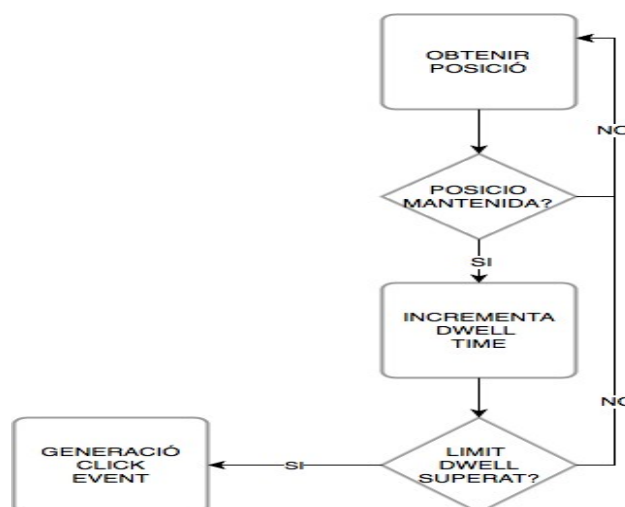


Figura 36: algorisme de click per *dwell-time*

## 4.7 Funcionalitats addicionals.

### 4.7.1. Obtenció del valor exposició de l'entorn.

Les condicions d'il·luminació de l'entorn són determinants en l'èxit de l'algorisme de seguiment. Condicions d'il·luminació intensa o insuficient, el rendiment de l'algorisme disminueix a no poder acumular prou punts característics de la cara suficient per portar a terme l'algorisme de seguiment amb èxit. Aquesta circumstància pot ser deguda a la naturalesa de la il·luminació o per la influència de factors externs, com ombres, oclusió de parts de la cara o reflexos de certs objectes.

Per la naturalesa d'aquest tipus d'interfícies és important proporcionar a l'usuari informació suficient que li permeti determinar l'estat de l'algorisme. Aquesta informació (*feedback*) pot ser utilitzada pels desenvolupadors d'aplicacions per informar l'usuari de les condicions de l'entorn.

El procediment de determinació del valor d'exposició es basa en la categorització del punt de màxima acumulació de l'histograma del canal de saturació en la conversió del *frame* a l'espai de color HSV. La Figura 37 il·lustra els passos del càlcul de l'histograma del valor d'exposició.

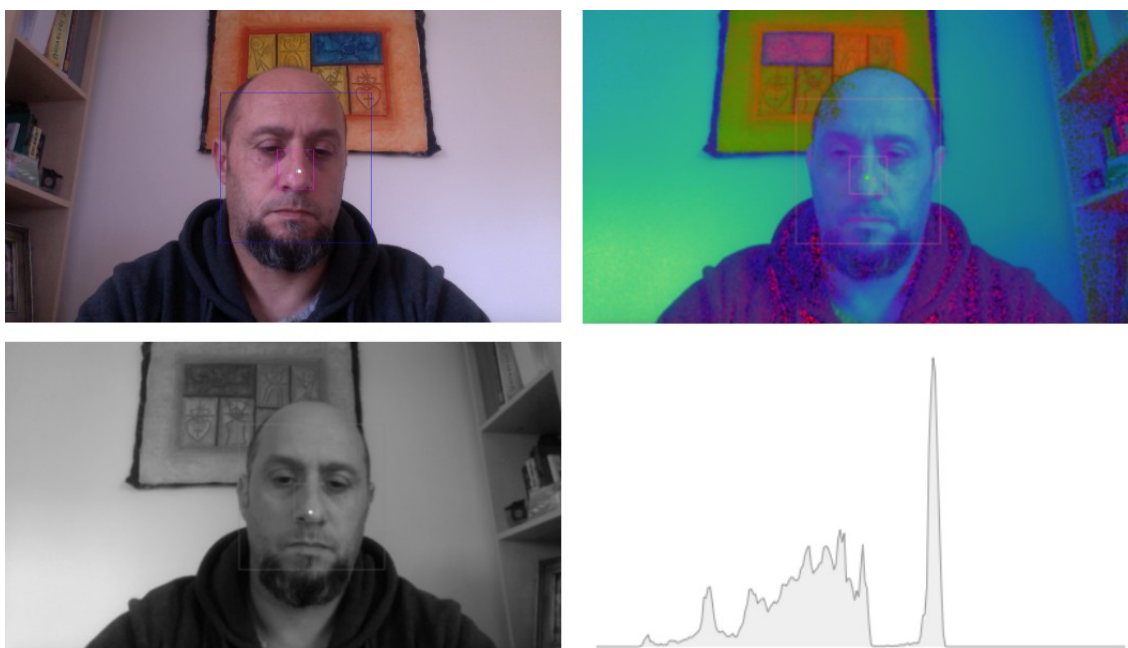


Figura 37: frame RGB, HSV, canal S i histograma resultant

La Figura 38 resumeix el procés d'estimació del valor d'exposició.

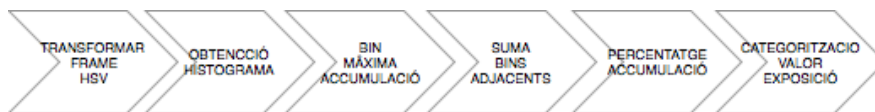


Figura 38: algorisme del valor d'exposició

El valor del grau d'exposició s'obté categoritzant la divisió de màxima acumulació de l'histograma. Per a reforçar aquest valor es calcula el percentatge d'acumulació respecte al total de la suma de la divisió de màxima acumulació i de les dues divisions adjacents.

La Figura 39 mostra una representació de la categorització de l'histograma.

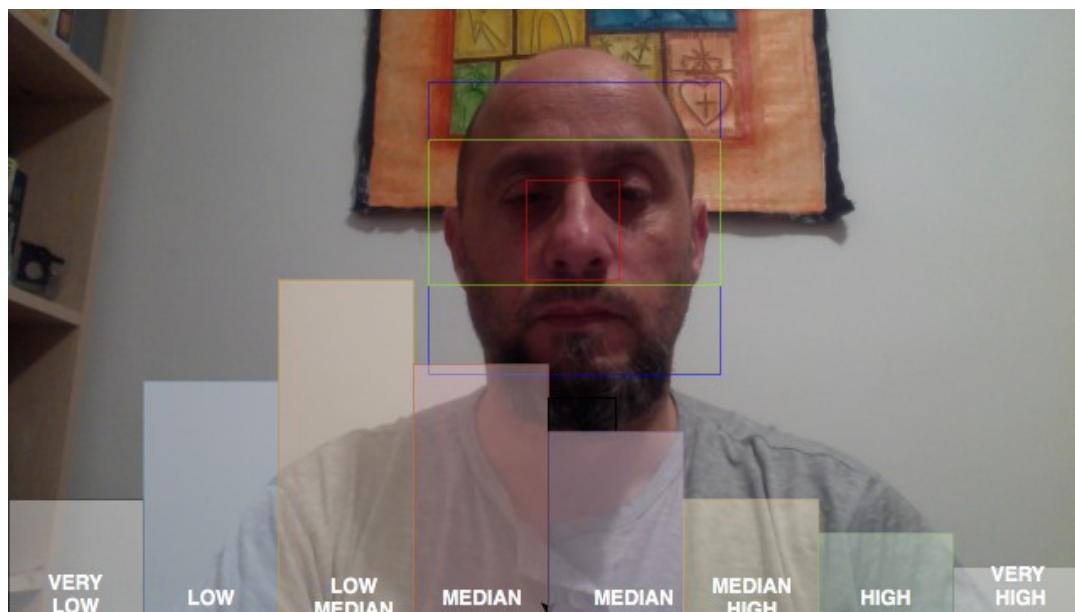


Figura 39: categorització del valor màxim d'exposició

El valor obtingut en el pas anterior es categoritza usant la següent escala de la Taula 8:

Taula 8: categorització de les condicions d'il·luminació

DESCRIPCIÓ	VALOR
Condicions de baixa il·luminació extrema	EXPOSITION_VERY_LOW
Condicions de baixa il·luminació	EXPOSITION_LOW
Condicions de il·luminació mitja-baixa	EXPOSITION_MEDIUM_LOW
Condicions de il·luminació mitja	EXPOSITION_MEDIUM
Condicions de il·luminació mitja-alta	EXPOSITION_MEDIUM_HIGH
Condicions de alta il·luminació alta	EXPOSITION_HIGH
Condicions de alta il·luminació extrema	EXPOSITION_VERY_HIGH

#### 4.7.2. Obtenció del valor posicionament i cobertura de l'usuari.

La posició de l'usuari respecte a la imatge obtinguda de la càmera és rellevant per al correcte funcionament de l'algorisme en la fase de transferència, on es produeix la transformació del moviment del punter de seguiment a les coordenades de pantalla.

L'efecte negatiu d'un mal posicionament de l'usuari es reduceix si:

- L'usuari es troba centrat respecte del *frame* (valor de desplaçament).
- L'usuari es troba a una distància pròxima de la càmera del dispositiu (valor de cobertura).

Amb l'objectiu d'ajudar els desenvolupadors a millorar la usabilitat de les seves aplicacions el sistema proporciona informació de *feedback* relativa al posicionament de l'usuari.

La Figura 40 mostra una representació de la divisió de les zones de desplaçament del punt central de la cara.

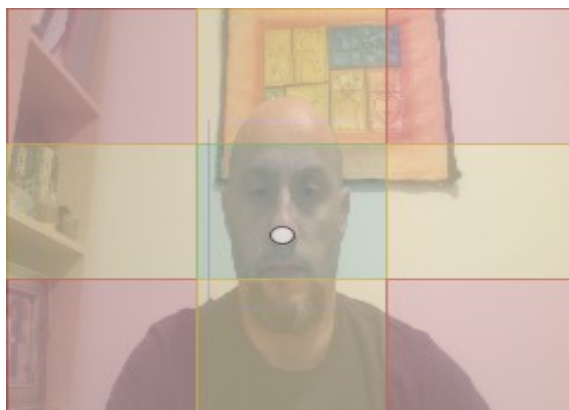


Figura 40: graella de zones de desplaçament

El valor de desplaçament es calcula a partir del valor del punt central de la cara detectada, aquest punt mitjà és vàlida respecte a la divisió de zones de la imatge anterior per a determinar el valor de posicionament de l'usuari.

El valor del desplaçament es categoritza segons el valor de la Taula 9:

Taula 9: categorització del posicionament de l'usuari

DESCRIPCIÓ	VALOR
Posicionament part central	DISPLACEMENT_SCORE_CENTER
Posicionament part superior-esquerra	DISPLACEMENT_SCORE_TOP_LEFT
Posicionament part central-esquerra	DISPLACEMENT_SCORE_LEFT
Posicionament part inferior-esquerra	DISPLACEMENT_SCORE_BOTTOM_LEFT
Posicionament part superior-dreta	DISPLACEMENT_SCORE_TOP_RIGHT,
Posicionament part central-dreta	DISPLACEMENT_SCORE_RIGHT

DESCRIPCIÓ	VALOR
Posicionament part inferior-dreta	DISPLACEMENT_SCORE_BOTTOM_RIGHT
Posicionament part central-superior	DISPLACEMENT_SCORE_TOP
Posicionament part central-inferior	DISPLACEMENT_SCORE_BOTTOM

El valor de cobertura es calcula a partir de les dimensions de la cara detectada en relació a les dimensions del *frame*.

La Figura 41 mostra una representació de la divisió en zones de cobertures de la cara.

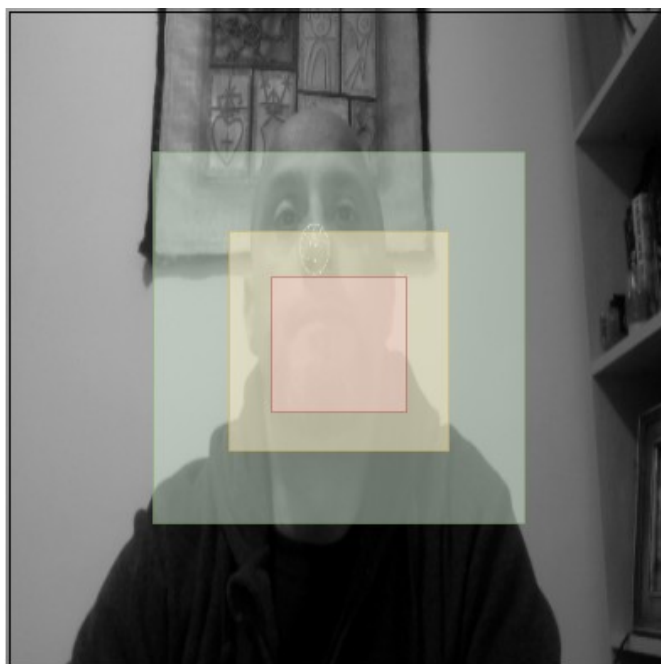


Figura 41: divisió en zones de cobertura

El valor de cobertura es categoritza segons el valors de la Taula 10:

Taula 10: categorització del grau de cobertura

DESCRIPCIÓ	VALOR
Grau de cobertura molt alt (per damunt del 60%)	COVERTURE_VERY_HIGHT
Grau de cobertura alt (en el rang [50%-60%])	COVERTURE_HIGHT
Grau de cobertura mitjà (en el rang [30%-50%])	COVERTURE_MEDIUM
Grau de cobertura baix (en el rang [20%-30%])	COVERTURE_LOW
Grau de cobertura molt baix (per sota el 20%)	COVERTURE_VERY_LOW

## 5. Aplicació de verificació

En aquesta secció es descriu la metodologia i accions realitzades a fi de verificar els productes generats en la fase d'implementació. Els detalls de l'estructura d'aquests productes i components poden ser consultats en els Capítol 3 i Capítol 4 d'aquest document.

Es verifica el compliment de les següents fites:

- Que la interfície de *head-tracking* permet el moviment d'un cursor de selecció virtual en l'àrea de la pantalla de l'aplicació mitjançant detecció i seguiment del nas, tal com es descriu en Capítol 4.
- Que el sistema és robust i tolerant a fallades puntuals en la detecció del nas.

La metodologia seguida per a la realització d'aquesta verificació consisteix en el desenvolupament d'una aplicació que actuï com a prova de concepte i que permeti validar el comportament del sistema en diversos dispositius amb diferents usuaris.

Per a incentivar als usuaris en l'ús de l'aplicació, aquesta ha estat dissenyada com un joc simple en el qual els usuaris han de moure un cursor virtual present en la pantalla amb l'objectiu de tocar diferents cercles de colors.

Aquests cercles de colors apareixen de forma aleatòria en diferents posicions de la pantalla i es mantenen actius durant un període de temps definit. Si l'usuari toca el cercle a temps es computa com un encert, en cas contrari com una fallada.

Les posicions en les quals poden aparèixer els cercles es defineix per la divisió de la pantalla en una graella de 11x11, configurant així un total de 121 posicions possibles. La Figura 42 mostra la graella de divisió de les posicions.



Figura 42: graella de posicions de l'aplicació



Durant el procés del cicle de joc l'aplicació recopila informació de diferents indicadors del rendiment de l'algorisme i del comportament de l'usuari en el joc. La Taula 11 mostra el llistat complet dels indicadors recollits per l'aplicació.

Taula 11: indicadors de l'aplicació de verificació

INDICADOR	DESCRIPCIÓ
HITS	Numero de moviments amb èxit
FAILS	Numero de moviments amb error
EXPOSITION	Valor modal de les lectures del valor de cobertura
COVERTURE	Valor modal de les lectures del valor de cobertura
DISPLACEMENT	Valor modal de les lectures del valor de cobertura
SELECTION TIMEOUT	Valor del temps de selecció mínim
SELECTION SIZE	Valor del radio de cercle de selecció
TOTAL FRAMES	Numero de <i>frames</i> processats
VALID FRAMES	Numero de <i>frames</i> processats amb èxit en la detecció
TOTAL TRACKING FRAMES	Numero de <i>frames</i> de tracking processats
VALID TRACKING FRAMES	Numero de <i>frames</i> de tracking processats amb èxit
TOTAL IDLE	Numero de estats IDLE
TOTAL CALIBRATING	Numero de estats CALIBRATING
TOTAL TRACKING	Numero de estats TRACKING
COMPUTE TIME	Temps total de processament
LOOP TIME	Temps mitjà de processament
FRAME MIN X FRAME MAX X FRAME MIN Y FRAME MAX Y	Rang de valors de la posició en coordenades de <i>frame</i>
TRACK MIN X TRACK MAX X TRACK MIN Y TRACK MAX Y	Rang de valors de la posició del punt de seguiment
SPEED MIN X SPEED MAX X SPEED MIN Y SPEED MAX Y	Rang de valors del desplaçament delta
SPEED_MAX	Valor màxim del valor del desplaçament delta
DISPLACEMENT MIN X DISPLACEMENT MAX X DISPLACEMENT MIN Y DISPLACEMENT MAX Y	Rang de valors de desplaçament
DISPLACEMENT_MAX	Valor màxim del valor del desplaçament

Amb la informació dels indicadors s'elaboren dues mètriques que ens ajuden a analitzar el rendiment de l'algorisme des del punt de vista dels objectius del projecte.

La mètrica de percentatge de frames de seguiment processats amb èxit («%tracking»), registra el nombre de frames que van ser processats amb èxit, durant la fase de seguiment. Aquest valor ens determina el nombre de frames en què va ocórrer pèrdues del seguiment, i ens permet analitzar el comportament de l'algorisme en el procés de seguiment d'una sessió d'usuari. El valor de la mètrica es calcula en base als indicadors *VALID\_TRACKING\_FRAMES* i *TOTAL\_TRACKING\_FRAMES*.

La mètrica de percentatge d'èxit en el cicle de joc («% hits»), registra l'èxit de l'usuari en el cicle de joc i ens permet analitzar la capacitat de l'usuari per a interaccionar amb els elements en la pantalla i per tant analitzar la capacitat de l'usuari de moure l'element punter per la superfície completa de l'aplicació. El valor de la mètrica es calcula en base als indicadors *HITS* i *FAILS*.

Al final del cicle de joc, la informació dels indicadors recollits es desa de forma persistent en l'àrea d'emmagatzematge extern del dispositiu.

Complementant la informació anterior recollida en el fitxer d'indicadors i mètriques, l'aplicació genera una imatge binària en la qual es mostra l'estat de selecció de cadascuna de les posicions per part de l'usuari en el cicle de joc. En cas d'èxit la secció corresponent a la cel·la de la graella s'acoloreix en color blanc, en cas va contrariar s'acoloreix de color negre. La imatge generada permet verificar el recorregut de la graella de selecció per part de l'usuari.

La Figura 43 mostra un exemple de la imatge binària obtinguda del comportament de l'usuari en el cicle de joc.

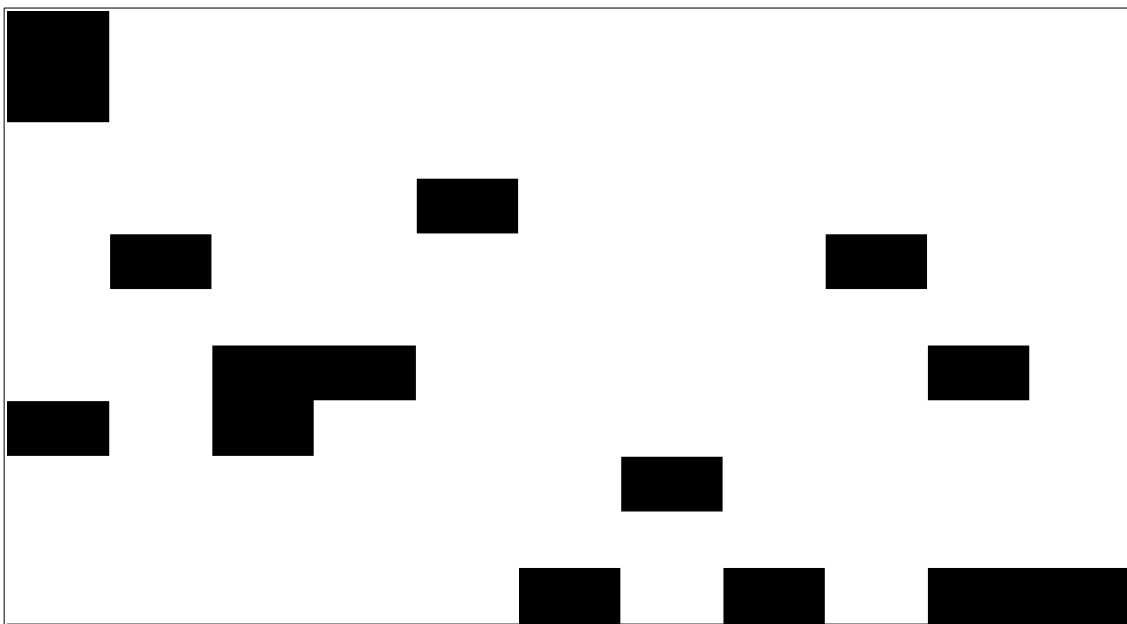


Figura 43: imatge binària del cicle de joc

## 5.1 Anàlisi dels resultats d'indicadors i mètriques

El conjunt global de totes mètriques recollides en els diferents dispositius pot ser explotada en un procés d'agregació estadística amb l'objectiu de l'extracció

d'informació rellevant que permeti millorar el rendiment de l'algorisme de visió i usabilitat.

Amb aquest objectiu es va modificar l'aplicació de prova de concepte per a afegir dues modalitats de joc:

- Mode de cobertura.
- Mode de precisió.

L'objectiu del mode de cobertura és validar que l'usuari té accés mitjançant el cursor virtual controlat per moviments del nas a totes les posicions de la pantalla. Per aquesta raó la dimensió del cercle s'ajusta a la dimensió màxima de la cel·la de la graella de divisió amb un temps de selecció és de 3-14 segons.

L'objectiu del mode de precisió és validar que l'usuari pot seleccionar zones petites de la pantalla. Per aquesta raó les dimensions dels cercles de selecció es configura com 50% de la dimensió de la cel·la de la graella de divisió i s'augmenta el temps de selecció a 8-19 segons.

La aplicació de prova de concepte ha estat instal·lada en diferents dispositius, les característiques dels quals es descriuen a continuació:

- Telèfon intel·ligent *Xiomi* model *Redmi 6* amb versió de android 8.1.0.
- Telèfon intel·ligent *Motorola* model *Moto G* amb versió de android 6.0.1.
- Telèfon intel·ligent *BQ* model *Aquarius C* amb version de android 7.1.0
- Tablet *Lenovo* model *TB-X304F* amb versió de android 7.1.1.
- Tablet *Nvidia* model *Shield* amb version de android 7.1.1.

Aquests dispositius han estat posats a la disposició d'una base de 8 usuaris amb un rang d'edat de 7 a 72 anys, cap d'aquests usuaris presenta cap grau de discapacitat.

En l'ús dels usuaris de l'aplicació no s'ha imposat cap condicions fixa de laboratori, sinó que es va permetre que l'usuari lliurement disposessin dels dispositius i els usessin al seu lliure criteri. L'objectiu d'aquest escenari és validar l'eficàcia del sistema head-tracking en entorns diversos amb diferents condicions d'il·luminació i posicionament de l'usuari.

L'estudi exhaustiu de les mètriques queda fora de l'abast del projecte pel fet que no es compta amb una mostra prou gran de persones dins del públic objectiu, en la qual s'han d'incloure persones amb discapacitat com principals beneficiaris de la interfície head-tracking.

Malgrat les limitacions de la mostra, l'anàlisi de la mostra obtinguda permet elaborar diverses hipòtesis interessants per a ser validades en un futur estudi formal.

La Figura 44 mostra el grau de correlació lineal de l'indicador del valor d'exposició i, les mètriques «% hits» y «% tracking». En relació a la mètrica percentatge de *frames* de seguiment processats amb èxit («% tracking»), observem que la mostra està molt esbiaixada per a extreure informació, en canvi en relació a la mètrica percentatge d'èxit en el cicle de joc («% hits») es pot observar una tendència. Aquesta tendència ens permeti elaborar la següent hipòtesi a validar, «a millor il·luminació de l'escena té com a conseqüència un millor rendiment de l'algorisme i els classificadors el que provoca una menor taxa de pèrdues de seguiment, el que té com a conseqüència major continuïtat en el procés de seguiment per a l'usuari i per tant més percentatge d'èxits».

La Figura 44 mostra la relació de correlació lineal de l'indicador del valor d'exposició i les mètriques.

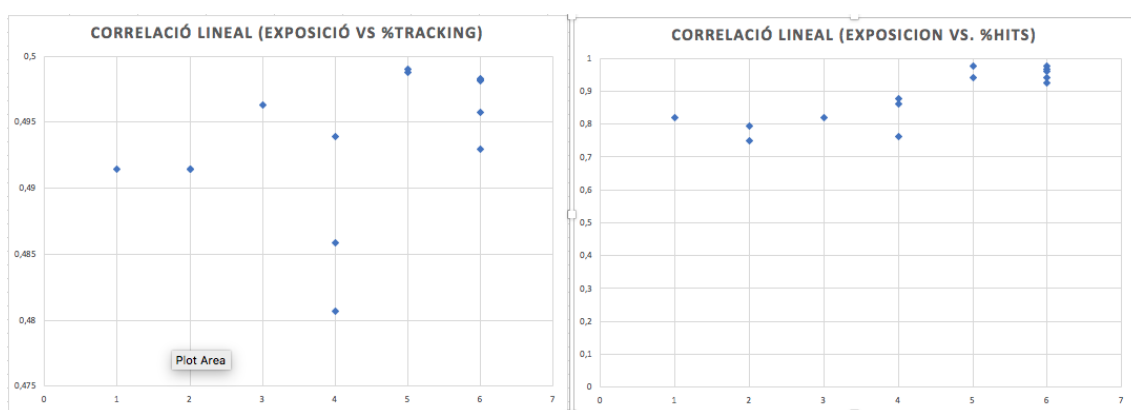


Figura 44: correlació de l'indicador d'exposició

Si analitzem la relació als indicadors de posicionament relatiu de l'usuari en el frame (Figura 45) en relació a la mètrica de percentatge d'èxit en el cicle de joc («% hits»), observem que es comporten com a variables independents, això ens permet elaborar la següent hipòtesi a validar, «l'usuari ajusta de manera intuïtiva la seva posició en el frame gràcies a la imatge de feedback a pantalla completa».

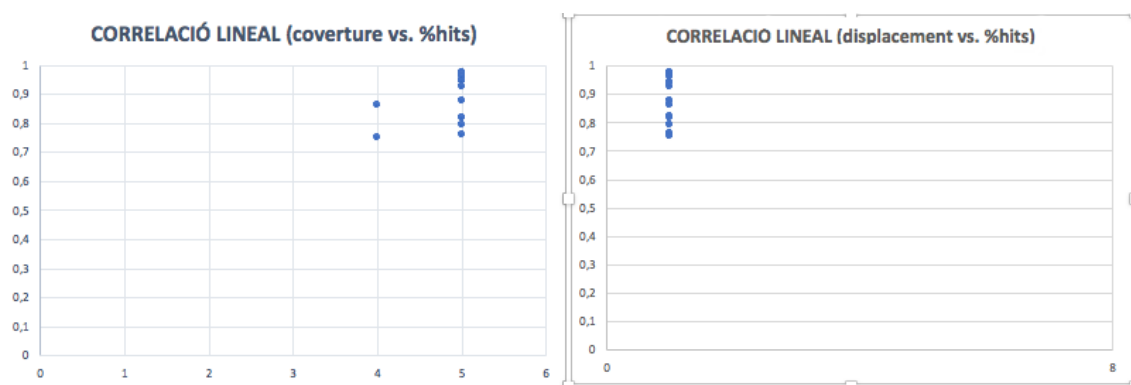


Figura 45: correlació de l'indicador de posicionament i cobertura

La Figura 46 mostra el grau de correlació lineal de l'indicador de valor de desplaçament delta, i, les mètriques «% hits» y «% tracking». El valor de desplaçament delta representa la intenció de moviment màxim de l'usuari

durant el cicle de joc. L'anàlisi de les gràfic demostra una tendència descendent de totes dues mètriques en funció del valor del desplaçament.

Aquesta tendència en totes dues gràfiques ens permet elaborar la següent hipòtesi a validar, «a major intenció per part de l'usuari de la realització de grans desplaçaments, el moviment del nas és major i per tant l'angle en relació a la càmera augmenta i com conseqüència es produeixen més pèrdues de seguiment a causa de fallades de detecció en el classificador frontal de cares».

La Figura 46 mostra la relació de correlació lineal de l'indicador del valor de desplaçament delta màxim i les mètriques.

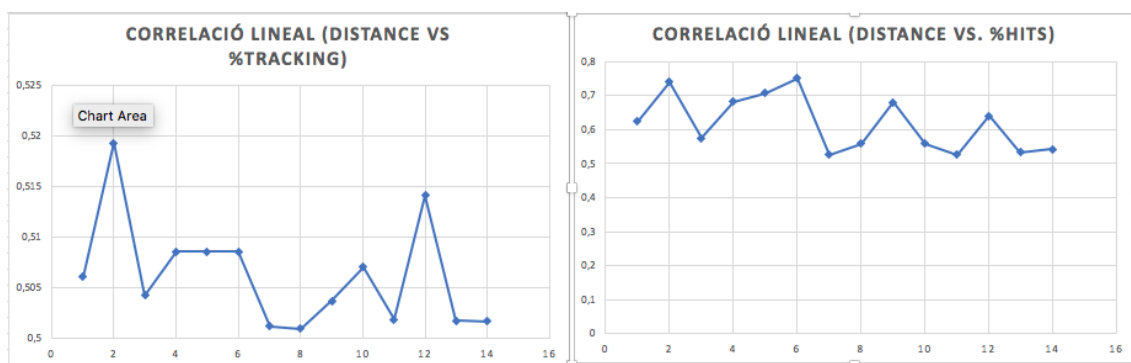


Figura 46: correlació de l'indicador desplaçament

L'agregació de les imatges binàries ens permet elaborar el percentatge d'èxit per a cada cel·la de la graella de selecció. Aquesta nova imatge generada pot ser utilitzada en la generació d'un mapa d'accessibilitat de les cel·les.

La Figura 47 mostra el mapa de calor resultant de totes les sessions dels usuaris registrades com a part del procés de validació, les zones en vermell intens es corresponen a les cel·les amb major percentatge d'èxits i les zones en b es correspon a les zones amb menor percentatge d'èxit.

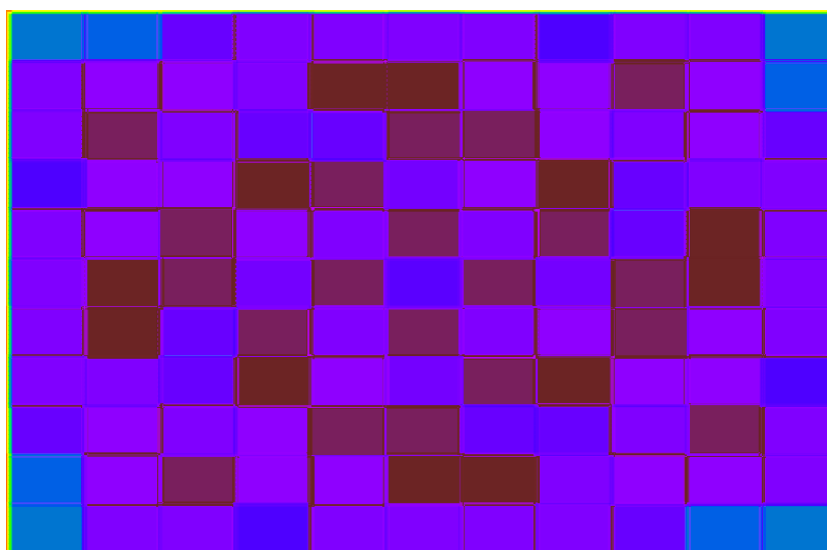


Figura 47: mapa de calor de les sessions dels usuaris

Tot i que la mostra és massa petita per extreure conclusions, la Figura 44 confirma que els punts més allunyats de la imatge corresponents als extrems de *frame* presenten menor puntuació. El motiu d'aquesta menor puntuació és la limitació de l'angle de detecció en el classificador frontal, tal com es va introduir en el capítol 4. L'estudi d'aquest comportament amb una mostra més àmplia d'usuaris ens ajudarà a confirmar la veracitat de l'afirmació.

## 6. Conclusions

En aquest projecte s'ha descrit el procés de disseny i implementació d'una interfície de visió per computador basada en el seguiment dels moviments del cap. La solució resultant és capaç de reconèixer i interpretar la posició del cap de l'usuari per a estimar la posició d'un element cursor en la pantalla del dispositiu mòbil.

Es va verificar que la interfície sigui intuïtiva d'usar per a l'usuari amb independència de les seves capacitats motrius i que proporciona de capacitats similars a les interfícies actuals.

La solució proposada es basa en un disseny modular. El disseny modular del sistema permet abstroure als desenvolupadors d'aplicacions dels detalls de l'algorisme de visió, facilitant la integració de la interfície en aplicacions noves i existents.

La metodologia de treball ha estat adequada per a permetre evolucionar el sistema sobre la base del disseny modular proposat. En una primera fase els esforços s'han centrat en els aspectes concrets de l'algorisme de visió, per a posteriorment enfocar-se en els detalls concrets de la plataforma.

Durant la fase de desenvolupat de l'algorisme de visió, s'han trobat amb problemes i dificultats, inicialment no previstes, que han estat necessari afrontar. En condicions reals d'execució les tècniques de visió i els classificadors presenten errors de detecció, aquests errors provoquen que el processament del *frame* no es realitzi amb èxit. Per a evitar que aquestes pèrdues de seguiment provoquen inestabilitat en l'algorisme de visió cal que s'implementin mecanismes de tolerància a fallades. L'estratègia de finestra de *frames* implementada en aquest projecte és un mecanisme eficaç per a dotar de robustesa a l'algorisme de visió.

S'ha constatat que les condicions de la il·luminació de l'entorn i el posicionament de l'usuari són factors determinants per a l'èxit del procés de seguiment, per aquesta raó s'han implementat funcionalitats addicionals per a informar la capa d'aplicació de l'estat de l'estat d'aquests dos factors.

Les capacitats motrius de tots els usuaris no són iguals, per aquesta raó és necessari la implementació d'un mecanisme d'ajust dinàmic de la transformació de l'espai de coordenades del *frame* a l'espai de coordenades de la pantalla. Aquest mecanisme ajusta la transformació de coordenades en els diversos eixos i direccions tenint en compte la capacitat motriu de l'usuari.

S'ha desenvolupat una aplicació de prova de concepte per a verificar que la solució compleix amb els objectius i requisits funcionals establerts inicialment. L'aplicació permet comprovar que l'usuari és capaç de moure de forma fluida un punter virtual per la pantalla de l'aplicació i que el sistema és estable.

A partir de l'aplicació de prova de concepte, s'ha elaborat un escenari d'ús, per a permetre l'obtenció d'indicadors i mètriques del rendiment de l'usuari i del comportament de l'usuari. L'escenari proposat es configura com un joc, en el qual els usuaris han de tocar cercles de color que apareixen de forma aleatòria en la pantalla. L'aplicació permet dues modalitats de joc:

- En la modalitat de cobertura es verifica que l'usuari és capaç de moure el cursor de forma fluida en la superfície de la pantalla.
- En la modalitat de precisió es verifica que l'usuari pot seleccionar objectes de menor grandària en la pantalla.

L'aplicació va ser instal·lada i posada a la disposició d'una base de 8 usuaris amb rangs d'edat de 7-72 anys. A causa de la grandària reduïda de la mostra i sense la pretensió de ser un estudi formal, els resultats obtinguts han permès l'elaboració de diverses hipòtesis sobre les quals iniciar un procés que derivi en un estudi formal.

El projecte m'ha permès adquirir coneixements de l'àrea de visió per computador i constatar les dificultats d'aquesta àrea de coneixement concreta. Com a conseqüència d'aquestes dificultats, he hagut d'introduir petites desviacions i correccions en la planificació inicial. Malgrat que aquests canvis en la planificació s'han pogut complir amb totes les fites establertes en la temporització.

Com a línia de treball futur s'estableix la col·laboració amb associacions de persones discapacitades i empreses de desenvolupament interessades en la promoció de l'ús de la interfície de visió. La col·laboració amb aquestes empreses permetrà el desenvolupament d'un estudi formal de l'ús i beneficis d'aquesta tecnologia per a les persones amb discapacitat.

Com a línia de treball futur existeix la necessitat de portar la implementació de la capa de plataforma de la interfície de visió a la plataforma IOS, pel fet que és juntament amb Android les plataformes dominants en la actualitat.



## 7. Glossari

- **Adaboost:** és un tipus d'algoritme d'aprenentatge automàtic on s'utilitzen múltiples classificadors febles per construir un algoritme d'aprenentatge més fort.
- **Computer vision (CV):** és un camp científic que tracta de com es poden fer ordinadors per obtenir una comprensió d'alt nivell d'imatges digitals o de vídeos.
- **Deep learning:** és un subconjunt de l'aprenentatge automàtic en intel·ligència artificial (IA) que té xarxes que imiten el funcionament del cervell humà en el processament de dades i la creació de patrons per utilitzar-los en la presa de decisions.
- **Dwell-time:** El període de temps en què un sistema o element d'un sistema es manté en un estat determinat, en el cas d'aquest projecte el període de temps què l'usuari manté una mateixa posició del punter virtual.
- **Eigenfaces:** és la tècnica de visió per computador per obtenir un conjunt de vectors propis quan s'utilitzen en el problema de la visió per computador del reconeixement de la cara humana
- **Eye-gazing:** és una àrea d'estudi que s'enfoca el determinar la posició de la pantalla en la qual l'usuari està centrant la seva atenció sobre la base de l'aplicació de tècnica de visió per computador.
- **Eye-gazing tracking:** és la aplicació de algorismes de eye-gazing per a la implementació d'interfícies de seguiment.
- **Feature point:** part rellevant de la imatge, tals com vèrtexs, cantonades, corbes o regions connectades. Aquestes característiques s'utilitzen com a punt de partida per a molts algorismes de visió per computadores.
- **Fisherfaces:** és l'aplicació de l'anàlisi del discriminant lineal per a obtenir una representació espacial d'un conjunt d'imatges de la cara, els vectors de base resultants que defineixen aquest espai es coneixen com fisherfaces.
- **Haar feature:** Les característiques de Haar són les funcions d'imatge digital utilitzades en el detector de cares.
- *head-tracking:* interfície d'usuari basada seguiment dels moviments del cap.
- **Heatmap image:** és una representació gràfica de dades on els valors individuals continguts en una matriu es representen com a valors d'una escala en forma de colors.
- **Human-computer interface (HCI):** La disciplina dedicada a dissenyar, avaluar i implementar sistemes informàtics interactius per a l'ús humà, i a estudiar els fenòmens relacionats més significatius.
- **Lambertian reflectance:** El model d'il·luminació que la propietat que defineix una superfície ideal reflectora difusa.
- **Machine learning:** És el camp de la intel·ligència artificial que es dedica a l'anàlisi i desenvolupo d'algorismes capaços de generalitzar

comportaments i patrons per a permetre que les compressió de la màquines evolucioni.

- **Motion tracking**: àrea de la visió per computador encarregada de determinar el moviment dels objectes en una seriosa d'imatges o vídeo.
- **Natural user interface (NUI)**: són aquelles interfícies d'usuari en les quals la interacció es realitza sense necessitat d'elements auxiliars. En el seu lloc la interacció és substituïa per moviments de parts del cos, detecció de postures o comandos de veu.
- **OpenCV framework**: és una biblioteca de programari de visió de computadora i d'aprenentatge automàtic de codi obert.
- **Optical flow**: és la detecció de patró del moviment aparent dels objectes, superfícies i vores en una escena.
- **Pitch rotation**: rotació al voltant de l'eix horitzontal X.
- **Pose estimation**: determinació de la posició i orientació d'un objecte en l'escena.
- **Pyramidal image**: tècnica de visió per computador de reducció de la imatge en una sèrie imatges sense pèrdua d'informació.
- **Template matching**: tècnica de visió per computador enfocada a trobar parts petites d'una imatge que coincideixin amb una imatge de plantilla.
- **Web Accessibility Initiative (WAI)**: Iniciativa promoguda pel consorci **World Wide Web (W3C)** per a millorar l'accessibilitat de la web per a les persones discapacitades.
- **World Wide Web (W3C)**: consorci internacional encarregar de desenvolupar i promocionar estàndards per al World Wide Web.
- **Widget**: Bloc de codi independent del programa que realitza una funció específica.
- **Wink gesture**: Detecció del gesto picada d'ullet per tècniques de visió per computador.
- **Xcode**: entorn de desenvolupament per a sistemes operatius **OSX**.
- **Yaw rotation**: rotació al voltant de l'eix vertical Y.

## 8. Bibliografia

- [1] HEWETT; BAECKER; CARD; CAREY; GASEN; MANTEI; PERLMAN; STRONG; VERPLANK (1996) «CHAPTER 2: Human-Computer Interaction». *ACM SIGCHI Curricula for Human-Computer Interaction* [en línia] <https://tinyurl.com/yy78enyr> [10/03/2019]
- [2] JOHN M. CARROLL (2009) «Conceptualizing a possible discipline of human-computer interaction». *Interacting with Computers, Volume 22, Issue 1* (pàg. 3–12).
- [3] JULIO CABERO ALMENARA (2008). *TICs para la igualdad: la brecha digital en la discapacidad*. [en línia]. [13/03/2019]
- [4] ESCANDELL, OLGA; BERMÚDEZ, ESCANDELL; SEVILLA, SOL; JOSÉ, LABORAL; CASTRO SÁNCHEZ, JOSE. (2014). «La brecha digital en las personas con discapacidad visual the digital divide among people with a visual disability». *International Journal of Developmental and Educational Psychology*,. 1. 214-9877489.
- [5] ORGANIZATCIÓ MUNDIAL DE LA SALUD (2011). *Informe mundial sobre la discapacidad*. [en línia] <http://alturl.com/a62po> [03/03/2019]
- [6] INSTITUT NACIONAL DE ESTADÍSTICA (2008). Encuesta de Discapacidad, Autonomía personal y situaciones de Dependencia (EDAD). [en línia]. <https://www.ine.es/prensa/np524.pdf>. [03/03/2019]
- [7] WEIYUAN LIU (2010). «Natural user interface- next mainstream product user interface». 2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design 1.
- [8] MURTHY R.; JADON G. (2009). «A review of vision based hand gestures recognition». *International Journal of Information Technology and Knowledge Management July-December 2009, Volume 2, No. 2* (pàg. 405-410).
- [9] MAGDALENA GONZÁLEZ; DÉBORA MULET; ELISA PEREZ; CARLOS SORIA; VICENTE MUT (2010). «Vision based interface: an alternative tool for children with cerebral palsy». *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*.
- [10] SIK CHANG, JAE; KIM, SUNG; KIM, HEE JOONG. (2005). «Vision-Based Interface for Integrated Home Entertainment System». *Pattern Recognition and Image Analysis* (pàg. 176-183).
- [11] CRISTINA MANRESA-YEE (2009). «Advanced and natural interaction system for motion-impaired users». Universitat de les Illes Balears [en línia]. <http://dspace.uib.es/xmlui/handle/10803/9412> [03/03/2019].
- [12] JAVIER VARONA; CRISTINA MANRESA-YEE; FRANCISCO J. PERALES (2005). «Hands-free vision-based interface for computer accessibility». *Journal of Network and Computer Applications Volume 31, Issue 4, November 2008* (pàg. 357-374).
- [13] M. BETKE; J. GIPS ; P. FLEMING (2002). «The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People With Severe Disabilities». *IEEE Transactions on Neural Systems and Rehabilitation Engineering, Volume: 10 , Issue: 1* (pàg. 1.10){
- [14] ANDERS HENRYSSON, MARK BILLINGHURST, AND MARK OLLILA (2005). «Face to Face Collaborative AR on Mobile Phones». *In Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '05)*.
- [15] MARIA FRANCESCA ROIG-MAIMÓ; SCOTT MACKENZIE; CRISTINA MANRESA-YEE; JAVIER VARONA (2018). «head-tracking Interfaces on Mobile Devices: Evaluation Using Fitts' Law and a New Multi-Directional Corner Task for Small Displays». *International Journal of Human-Computer Studies Volume 112* (pàg. 1-15).

- [16] MIGUEL BORDALLO LOPEZ; JARI HANNUKSELA; OLLI SILVÉN; LIXIN FAN (2012). «*head-tracking* virtual 3-D display for mobile devices». *Conference: Embedded Vision Workshop (CVPR12)*.
- [17] HEIKO DREWES; ALEXANDER DE LUCA; ALBRECHT SCHMIDT (2007). «Eye-gaze interaction for mobile phones». *Mobility '07 Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology* (pàg. 364-371)
- [18] TAKASHI NAGAMATSU, MICHIIYA YAMAMOTO, AND HIROSHI SATO (2010). «MobiGaze: development of a gaze interface for handheld mobile devices». In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*.
- [19] KYLE KRAFKA, ADITYA KHOSLA, PETR KELLNHOFER, HARINI KANNAN, SUCHENDRA BHANDARKAR, WOJCIECH MATUSIK, ANTONIO TORRALBA; (2016). «Eye Tracking for Everyone». *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016* (pàg. 2176-2184).
- [20] KAEHLER ADRIAN; BRADSKI GARY (2016). *Learning OpenCV 3*. United States of America. O'Reilly.
- [21] GONZALEZ RAFAEL; WOODS RICHARD (2008). *Digital Image Processing – Third Edition*. United States of America. Pearson International Edition.
- [22] IKEUCHI, KATSUSHI (2014). "Lambertian Reflectance". *Encyclopedia of Computer Vision*. Springer. (pàg. 441–443).
- [23] SCHIEBER FRANK (2000) «Photometry for Traffic Engineers Slideshow». *TRB 2000 Human Factors Workshop Photometry for Traffic Engineers* [en línia] <https://tinyurl.com/lyxubw3pc> [30/04/2019]
- [24] YOW CHOONG KIN; CIPOLLA ROBERTO (1997) . "Feature-based human face detection". *Image and Vision Computing* Volume 15, Issue 9 (pàg. 713-735).
- [25] Singh Sanjay Kr.; Chauhan D. S. ; Vatsa Mayank ; Singh Richa (2003) "A Robust Skin Color Based Face Detection Algorithm" 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010).
- [26] FUNDACIÓN VODAFONE ESPAÑA «EVA facial mouse». [en línia]. [https://play.google.com/store/apps/details?id=com.crea\\_si.eviacam.service](https://play.google.com/store/apps/details?id=com.crea_si.eviacam.service) [03/04/2019]
- [27] CHENG DANIEL; VERTEGAAL ROEL (2004). "An Eye for an Eye: A Performance Evaluation Comparison of the LC Technologies and Tobii Eye Trackers». *Human Media Lab, Queen's University, Canada*
- [28] ORIGIN INSTRUMENTS. «*Camera HeadMouse Nano*». [en línia] <http://orin.com/access/headmouse/>[03/04/2019]
- [29] DIVYA VENUGOPAL ; J AMUDHA ; C. JYOTSNA (2016). «Developing an application using eye tracker». *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*
- [30] VARIS AUTORS; «RGB Color space». Wikipedia [en línia] <http://alturl.com/oax32> [03/04/2019]
- [31] VARIS AUTORS; «HSV Color space». Wikipedia [en línia] <http://alturl.com/psr7g> [03/04/2019]
- [32] VARIS AUTORS; «YCbCr Color space». Wikipedia [en línia] <http://alturl.com/nre3c> [03/04/2019]
- [33] MING-HSUAN YANG; (2002). "Face Recognition Using Kernel Fisherfaces". United States Patent, Patent No.: US 7,054,468 B2
- [34] HESPANHA J.; KRIEGMAN D.; BELHUMEUR P. (1997). "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 19, no. 07 (pàg. 711-720)
- [35] JONES MICHAEL; VIOLA PAUL (2001). "Rapid object detection using a boosted cascade of simple feature". *IEEE Conf. on Computer Vision and Pattern Recognition* Kauai, Hawaii.

- [36] JONES MICHAEL; VIOLA PAUL (2003). "Multi-view Face Detection". [en línia] <https://urlzs.com/ZG81> [20/04/2019]
- [37] FREUND Y.; SCHAPIRE R. (2000) "A decision-theoretic generalization of on-line learning and an application to boosting". *Journal of Computer and System Sciences* 55(1) (pàg. 119–139)
- [38] SIONG; MOKRI; HUSSAIN ; IBRAHIM; MUSTAFA (2009) «Motion detection using Lucas Kanade algorithm and application enhancement». 2009 International Conference on Electrical Engineering and Informatics
- [39] JOEL LE ROUX (2003) "An introduction to Kalman Filtering :Probabilistic and Deterministic Approaches", University of Nice
- [40] SCHERR; LOMMATSCH; HOOSON; KAISER; "Underwater Navigation System" [en línia] <http://alturl.com/x3ptm>
- [41] QUINTERO FORERO; ADRIANA (2016). «*Extraction of morphological features of human face from images*». Documento trabajo de grado UMNG.
- [42] COMANICIU D; MEER P. (2002). «Mean Shift: A Robust Approach Toward Feature Space Analysis». *IEEE Transactions on Pattern Analysis & Machine Intelligence May 2002* (pàg. 603-619)
- [43] NOUAR D; ALI G.; RAPHAEL C. (2006). «Improved Object Tracking With Camshift Algorithm».
- [44] BOUGET JEAN-YVES (2002). «Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker». Intel Corporation [en línia] <https://tinyurl.com/y38ofhgm> [03/03/2019].