UNIVERSITAT DE BARCELONA

Treball final de grau

GRAU D'ENGINYIERIA INFORMÀTICA

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

---

# SISTEMA DE RE-IDENTIFICACIÓ DE PERSONES BASAT EN CARACTERÍSTIQUES FACIALS I CONTEXTUALS

---


Autor: Alexandre COLA SABORIT

| | |
|---|---|
| Directora: | Estefanía Talavera M. |
| Co-directora: | Dr. Petia Radeva |
| Realitzat a: | Dept. de Matemàtiques i Informàtica |
| Barcelona, | 2 de febrer de 2018 |

# Resum

En el 2003 Gordon Bell [1] va crear un diari de vida, el qual registrava de forma automàtica la vida quotidiana per mitjà de càmeres portables (wearable cameras), on es recopilaven imatges des del punt de vista de primera persona. Una pregunta natural que apareix és: com podem recuperar totes les imatges on apareix la mateixa persona? El concepte de re-identificació [4] s'utilitza per a tornar a detectar persones que s'havien localitzat anteriorment en altres imatges. Amb cada detecció s'acumula més informació per facilitar noves deteccions.

En aquest treball es proposa la re-identificació de persones per mitjà d'aspectes facials i contextuals a través d'imatges egocèntriques. A més, s'implementa un algoritme de distorsió del rostre de les persones que apareixen en les imatges, per tal de garantir la privacitat d'aquestes.

A pesar de l'extensa literatura sobre re-identificació o *person re-identification*, en aquest treball expliquem més en detall dos models recents enfocats en imatges egocèntriques es dir adquirides amb càmeres portables. Una d'elles és un detector d'interaccions socials [7] que construeix un mapa bidimensional de les persones amb qui interactua. Amb la teoria de F-formació [13] es determina si hi ha interacció social, analitzant la seqüència d'imatges. L'altra aplicació caracteritza l'estil social [8] del portador de la càmera, per mitjà de la distància entre les persones, les diferents posicions del cap de les persones que apareixen a les imatges i l'expressió social. La suma d'aquestes característiques concreta quin tipus d'interacció social té el portador de la càmera amb les persones de les imatges.

L'algoritme de privacitat s'implementa primer localitzant i després distorsionant la cara detectada. Per tal de localitzar les cares que apareixen en les imatges egocèntriques s'utilitza un algorisme anomenat Viola & Jones [18], el qual està integrat en les llibreries de OpenCV [17].

Una vegada localitzada la cara, s'utilitza informació d'aquesta i del seu context per a dur a terme l'agrupament amb altres cares detectades. Gràcies a la localització del rostre en les imatges, es retalla; en el retall de la cara també s'incorpora la part superior del pit que s'utilitzarà per extreure els valors de la roba que porta la persona, així es pot analitzar el context relacionat amb la persona. Amb la imatge del rostre s'extreu un vector característic, per mitjà de la llibreria OpenFace [20], el qual està compost de 128 valors que representen la cara de la persona que apareix.

Amb tots els vectors característics dels rostres es comproven diversos sistemes d'agrupament, concretament Jeràrquic, MeanShift i el Spectral clustering [21] [27] [22]. Es comprova l'eficàcia dels sistemes d'agrupament per organitzar els rostres per clústers, on es proposa que cada clúster representi una persona concreta. Per millorar els resultats dels clústers, es proposa utilitzar el Coeficient de Correlació de Pearson [23]. Amb aquest coeficient es comprova la consistència dels clústers i amb aquest valor es verifica que la precisió de l'agrupament augmenta de manera considerable, donant resultats entre el 80% i 100% de precisió, es a dir, repartint les imatges de les persones en un clúster per persona.

Es proposa utilitzar informació del context de la persona detectada incorporant valors descriptius de la seva roba. Es localitza la zona de la roba per mitjà de les proporcions del Golden Ratio [28] on es determinen segons les dimensions de la cara quines proporcions té el cos per poder extreure la part superior del pectoral i així ubicar la zona de la roba. Per a recopilar els millors valors de la roba es comproven diferents espais de color (YCrCb, HSV, BGR i Gray). Per a extreure sols valors representatius de la roba, es detecta la pell de la persona, per mitjà dels colors detectats en la zona del rostre. Es proposa utilitzar l'espai de colors HSV, el qual descartant la component de la lluminositat aporta uns valors concrets quals en bastants casos remarquen la zona de la pell. Una vegada detectada la composició de la pell, es descarten els seus píxels, sense els quals s'obtenen els valors més concrets de la roba que porta la persona. Amb els valors que s'extreuen de la roba, es crea el seu vector característic, que es concatena amb el vector característic del rostre. Es proposa la incorporació d'aquestes dades per millorar l'agrupament dels sistemes jeràrquic, MeanShift i el Spectral clustering. També es comprova la distància de Bhattacharyya [24], utilitzada per la comparativa entre vectors característic de la roba de les persones detectades en les imatges egocèntriques.

# Agraïments

Expressar agraïment pel temps dedicat i sobretot la paciència de les directores d'aquest projecte, la Srta. Estefania Talavera i la Dra. Petia Radeva, que han estat en tot moment donant suport perquè aquest projecte pogués culminar. El temps ha estat ajustat, però durant el transcurs d'aquest treball s'ha agraït molt la dedicació i constància de les directores, on s'ha notat la seua implicació.

En tot moment s'ha vist una denotada professionalitat, la qual ha perdurat fins a l'últim moment. I per finalitzar, sols queda desitjar assolir nous reptes amb la mateixa voluntat i perseverança.

# Índex

# 1 Introducció

Actualment, hi ha moltes aplicacions que creen un diari de la nostra vida social, així com activitats, reunions, llocs habituals inclús recomanació de música. Per exemple, existeixen les aplicacions d'interaccions socials de *Facebook* (`https://facebook.com`) i *Twitter* (`https://twitter.com/`), les quals són les més conegudes a escala mundial. Aquestes aplicacions fan un registre públic de la vida quotidiana de l'usuari, però per portar aquest registre l'usuari ha de tenir una constant dedicació, ja que la informació no es registra automàticament. Un altre exemple seria l'aplicació de *Spotify* (`https://www.spotify.com`), la qual està dedicada a escoltar íntegrament música, però contempla la selecció dels temes musicals més habituals i la seva selecció de forma automàtica segons l'ús de l'usuari.

En el 2003, apareixia el concepte *Lifelogging* [1] que es definia com la realització d'un registre de la vida de l'usuari. En aquest projecte, volem estudiar el concepte de *Lifelogging* [1] relacionat amb el factor social, és a dir, amb qui ens relacionem. Per a la realització d'aquest projecte, s'utilitza una càmera portable (wearable cameras), la qual serà una **Narrative Clip 2-3fpm**, de 5 megapíxels de definició, d'on s'analitzaran les imatges gravades. Aquesta càmera enregistrarà de forma **automàtica** i continuada la vida del seu portador sense que aquest hagi d'interactuar amb la càmera. Amb aquest projecte es vol crear un sistema que realitzi una anàlisi social de la vida de l'usuari. El nostre objectiu és reconèixer les persones amb les que es relaciona l'usuari de la càmera per mitjà de les imatges egocèntriques que s'enregistren.

Tot l'exposat té una aplicació directa amb persones que sofreixen malalties mentals com l'Alzheimer. L'objectiu és que puguin recuperar el seus moments socials captats amb la càmera i així entrenar la seva memòria i les seves habilitats cognitives, per a alentir el deteriorament de les seves actituds i/o aptituds socials i cognitius.

## 1.1 Què és Lifelogging?

Lifelogging [1] és un registre de la vida, el qual es pot fer per mitjà de càmeres portables (wearable cameras) i/o amb algun altre dispositiu. La traducció d'aquest és la creació d'un diari de vida. Quan apareixen les càmeres, dóna lloc a col·leccionar les imatges que descriuen la vida d'una persona. Les esmentades col·leccions es coneixen com a egocèntric photo-streams.

Amb les col·leccions d'imatges es crea un extens registre de la vida quotidiana d'una persona. Gràcies a l'avanç de les noves tecnologies de la comunicació i emmagatzematge de la informació, es pot estudiar aquesta gran quantitat de dades.

L'enginyer informàtic, Gordon Bell [1] és un dels destacats pioners del concepte Lifelogging. Va començar aquesta idea quan anava transportant incòmodament papers i fotografies d'un lloc a l'altre, pel que va optar per guardar-los digitalment. Al començar a emmagatzemar tota aquesta informació, va incrementar la seva necessitat de guardar més dades. El volum de la informació va creixa de tal manera

que va portar al Sr. Bell a confeccionar un software perquè guardes automàticament tota la seva vida.

Per a desenvolupar aquest software (Fig. 1) es va ajuntar amb dos investigadors de Microsoft: Jim Gemell i Roger Lueder. Amb aquesta unió va sorgir la tendència aquí esmentada com a *Lifelogging*, la qual constava en registrar digitalment totes les activitats d'una persona.

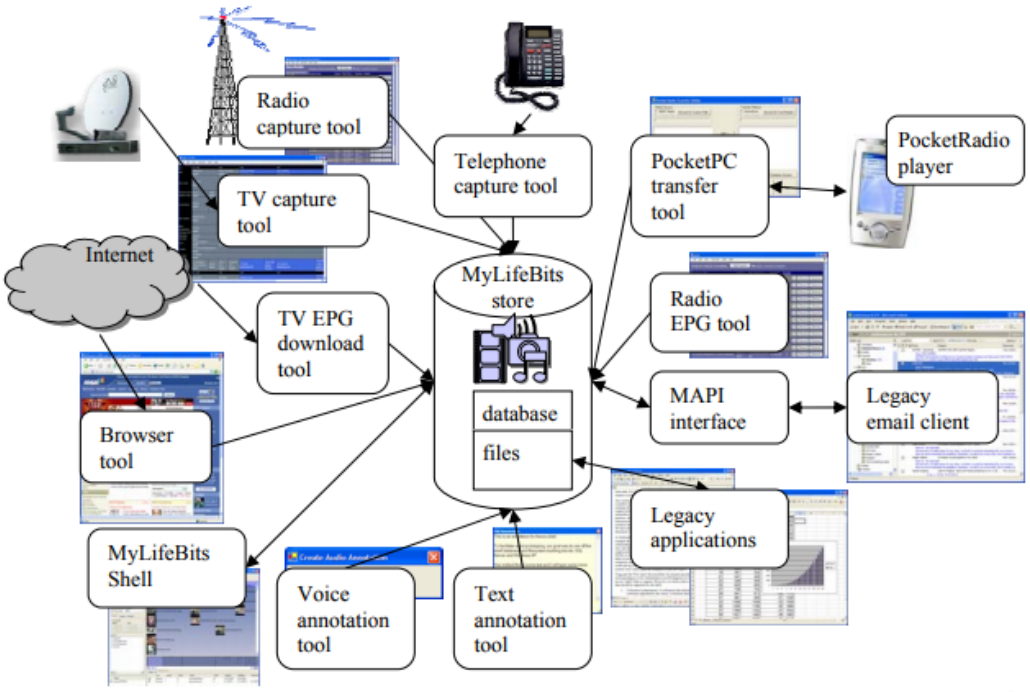


Figura 1: Esquema d'aplicació de Lifelogging [1]

Aquest concepte ha estat molt explotat i un dels exemples més actuals és l'aparició de les Google Glasse de l'empresa Google. Per mitjà d'unes ulleres, es va fent un registre constant de tot els que perceben els ulls amb la incorporació de realitat augmentada per a més informació. També van sorgir altres elements per enregistrar la informació de tota persona, com és la incorporació d'una càmera de dimensions reduïdes, com una petita xapa i sense botons, coneguda com a MEMOTO (avui en dia, Narrative). Aquesta petita càmera pren fotografies de manera automàtica donant referències geogràfiques de les imatges. També realitza captures cada 30 segons, és a dir, unes 10.000 imatges per cada 12h al dia en una setmana.

## 1.2 Egocentric Vision i wearable cameras

Egocentric Vision o first-person és una branca d'estudi de la visió artificial que implica l'anàlisi d'imatges i vídeos capturats per mitjà d'una càmera portàtil (wearable camera). Aquestes càmeres es col·loquen normalment en el front, en el cap o el pit de la persona que la porta (veure Fig. 3). Normalment enfoca a l'àrea de visió de la persona. Les dades capturades pel dispositiu ofereixen una perspectiva de l'es-

cena que l'usuari enfoca. En aquest treball, pretenem extreure informació d'aquest context per a millorar la classificació de persones. Per aquest motiu, inclourem informació de la roba que porta la persona amb què s'interactua.


Figura 2: Primeres càmeres portàtils (Wearable cameras)

La idea d'utilitzar una càmera portàtil (wearable camera) per recopilar dades visuals des d'una perspectiva en primera persona (Egocentric Vision o first-person) es remunta als anys 70 (Fig. 2), quan Steve Mann va inventar *Eye Glass*. Es tractava d'un dispositiu electrònic que captura les imatges i les representa en una pantalla de televisor. Però sols després d'entrar al mercat en el 2006 la *SenseCam* de Microsoft, vas ser quan es van utilitzar per primera vegada les càmeres portàtils per a treballs en la investigació a gran escala.

L'interès de la comunitat de la visió per computació en el paradigma egocèntric ha anat en augment, però aquest interès va sorgir a partir del 2010. Aquest interès apareixia gràcies a la creixent i innovadora tecnologia dels dispositius portables.

Figura 3: Càmeres portàtils : SenseCam i Narrative Clip

Per aquest projecte s'utilitzarà una càmera portable (wearable camera) **Narrative Clip 2-3fpm** de 5 megapíxels de definició, la qual es mostra a la dreta de la figura Fig. 3. Aquesta càmera és capaç d'adquirir fins 2000 imatges per dia amb una autonomia de fins 2 dies. És molt fàcil d'usar-la ja que no té cap botó, dispara quan hi ha llum i una vegada connectada a l'ordinador descarrega les imatges de forma automàtica i carrega la seva bateria. Lúltima versió de Narrative (Clip 2) també disposa amb l'opció de descarregar les imatges per la xarxa sense fil (wi-fi).

## 1.3 Person Reidentification

El concepte de *person reidentification* (re-identificació d'una persona) apareixia per ajudar als algorismes de les càmeres de vídeo-vigilància [10] [3] [4]. Els mètodes utilitzats es basaven en l'anàlisi i comparació d'informació extreta dels cossos dels vianants.

El propòsit de la re-identificació d'una persona és detectar a una persona d'interès. Quan aquest va sorgir, els algoritmes eren més simples i tenien una avaluació a petita escala que s'anava informant periòdicament. Els últims anys han aparegut conjunts de dades a gran escala i sistemes per *Deep Learning*, que fan ús de grans volums de dades.

Es consideren diferents tasques, on es classifiquen la majoria dels mètodes d'identificació en dues classes, els basats en imatges i els basats en vídeos. En aquest projecte, sols es tracta el sistema basat amb imatges. A més, es discuteixen dues noves tasques d'identificació que estan més pròximes per l'aplicació en el món real, és a dir, la identificació completa per mitjà de l'explotació d'imatges individuals de la persona (single shot) i la re-identificació ràpida realitzada per mitjà d'extenses galeries d'imatges [4].

L'objectiu de person reidentification és identificar a una persona quan aquest es torna a presentar, però en una nova vista, és a dir en un nou context. Aquest concepte es podria aplicar fàcilment en aplicacions de vídeo-vigilància per a seguretat pública i/o privada.

El concepte de person reidentification es pot tractar per característiques facials, però s'ha de tenir en compte que l'objectiu final és la identificació d'una persona d'interès. Es pot aconseguir el mateix objectiu utilitzar altres mètriques com per exemple les dimensions del cos, tal com tracte l'article *Partial Person Re-identification* [3], on es proposa la identificació d'una persona per mitjà d'una part del cos.

En aquest projecte es treballa amb el person reidentification d'imatges per mitjà de característiques facials ajudades amb altres elements contextuals, com és la roba.

## 1.4 Motivació

La motivació d'aquest treball és poder aportar eines per a l'estudi, i posterior anàlisi i millora de la vida social de la gent.

Qualsevol persona pot fer ús d'aquestes eines per tal de poder analitzar o millorar el seu entorn social, així com tenir un registre de la seva vida quotidiana. També, si l'usuari té o vol tenir una notòria vida social a les xarxes, podria beneficiar-se d'aquestes eines, ja que automatitzaria les seves publicacions. Això facilitaria tenir una vida social a la xarxa més activa, on publicar les seves imatges de manera automàtica i sense interaccions podria reduir el temps que dedica en anar generant publicacions.

Però a part de tenir un ús merament social, aquestes eines es poden explotar

amb finalitats mèdiques. Hi han persones que tenen dificultats, com per exemple els afectats d'Alzheimer, ja que si no mantenen un entrenament constant de les seves habilitats cognitives, aquestes es perden de manera més accelerada. Amb aquestes eines poden tenir un entrenament més constant del seu record i també poder recordar a determinades persones per no ser fruit de cap engany al trobar-se desorientats per culpa de la seva afectació. Per exemple, tota la informació recopilada es podria organitzar i preparar per ser utilitzada en una sèrie de jocs, els quals estarien dissenyats per mantenir les habilitats, actituds i aptituds socials per mitjà de l'exercitació mental reiterada.

També es pot contemplar l'ús d'aquestes eines com un ajut a la vida quotidiana de les persones que tenen alguna deficiència visual, com pot ser una persona cega o amb una greu minoració de la seva capacitat visual. Aquestes eines poden tenir un recopilatori de totes les persones que han vist i, fins i tot, es podria implementar en un audiòfon amb càmera, perquè l'usuari del dispositiu sentis el nom de la persona amb qui parla, i potser poder obtenir més informació com l'aniversari, estatura, possibles deficiències, etc.

## 1.5 Objectiu

Per concloure la introducció, l'objectiu d'aquest projecte és obtenir un conjunt d'algorisme que donin com a resultat una òptima classificació de les imatges utilitzades pels usuaris portadores de les càmeres portables. Gràcies a la classificació i anàlisis de les imatges, es podrà obtenir informació diversa dels usuaris de les càmeres portables.

Els algorismes es pretenen que siguin òptims amb precisió, ja que pel tipus d'informació que es vol processar, és valorar molt més la precisió de les imatges que els descarts d'aquestes. Per aquest motiu, es tindrà més insistència amb els falsos positius que els falsos negatius, ja que tal com s'ha dit, es valorarà per sobre de tot la precisió dels algorismes per a poder fer una re-identificació amb garanties.

## 1.6 Estructura de la memoria

Aquest treball es distribueix en 5 apartats, els quals són:

**Estat de l'art**: En aquest apartat es detallaran dues aplicacions vinculades amb la detecció de persones i les seves interaccions. Una detallarà les interaccions que realitza un usuari amb altres usuaris del seu voltant i l'altre detallarà el tipus d'interacció que es realitza amb determinades persones.

**Re-Identificatió de Persones**: En aquest apartat s'exposaran els mètodes proposats envers la distorsió de les cares per privatitzar-les i l'agrupament d'aquestes per realitzar la re-identificació. També s'explicarà el mètode proposat per obtenir més informació incorporant la roba de la persona detectada per tal d'afegir informació contextual per l'agrupament.

**Configuració Experimental**: En aquest apartat s'exposaran les dades que

s'utilitzaran per als experiments i es detallaran com es faran aquests. S'explicaran els sistemes de validació que s'ha utilitzat i les mesures amb les quals han portat a terme els diferents experiments.

**Resultats**: En aquest apartat es mostraran els resultats dels experiments proposats en l'apartat de Configuració Experimental. També s'analitzaran les dades obtingudes per a discutir els resultats.

**Conclusions i treball futur**: En aquest apartat s'exposaran les dades més rellevants del treball per tal de ser discutits i donar idees finals contrastades amb els experiments realitzats en altres apartats. També es vol donar una possible orientació de futur envers les dades i idees obtingudes.

# 2 Estat de l'art

En aquest apartat s'explicaran dues aplicacions dedicades a analitzar les interaccions socials d'un usuari, el qual porta una càmera portable (wearable camera). La primera aplicació està dedicada a la detecció d'interaccions socials [7] mentre que la segona aplicació estarà dedicada a caracteritza quin estil social [8] té la interacció, si n'hi ha. I l'última esta dedicada a la agrupació de les persones agrupant aspectes [19].

En la detecció d'interaccions socials s'explicarà l'anàlisi de les imatges obtingudes, on la posició i orientació del rostre de les persones podran determinar si hi ha interacció o no. Per altra banda la caracterització d'estil social analitza diversos valors del rostre com posició, inclinació i/o orientació del cap, on també se li suma expressió facial, per poder classificar quin tipus d'interacció (si n'hi ha), realitzem amb determinades persones.

Totes tres aplicacions parteixen del mateix punt de vista egocèntric, ja que els usuaris porten una càmera (wearable camera) que enfoca a la seva part frontal per enregistrar amb qui interactuen.

## 2.1 Detector d'interaccions socials

La detecció de les interaccions socials, es va introduir recentment a [7]. Aquesta aplicació analitza les imatges capturades per mitjà d'una càmera portable, la qual enregistra a baixa freqüència fotogrames i es porta per una persona durant tot el dia. L'objectiu d'analitzar les imatges capturades és poder concretar quan l'usuari s'involucra en una interacció social.

Figura 4: Representació bidimensional de l'entorn [7]

Per tal de determinar si l'usuari està interactuant amb una persona, es planteja el concepte psicològic de *F-formation*, el qual explota la distància i orientació de les persones respecte a l'usuari i que interactuen amb aquest. El resultat és un Patró d'interacció en una seqüència representada en una sèrie temporal bidimensional (Fig. 4) que correspon a l'evolució temporal de les característiques de distància i orientació al llarg del temps.

Per tal de classificar tota aquesta informació s'utilitza una xarxa neuronal recurrent basada en memòria a curt i llarg termini (LSTM) [5]. S'ha experimentat amb

un conjunt de 30.000 imatges amb resultats bastant prometedors.

Diferents treballs s'han dedicat al reconeixement automàtic i comprensió de les interaccions socials per mitjà de vídeos que combinen la informació d'altres dispositius com Bluetooth i infrarojos, [11], [12]. Però la definició d'una interacció social depèn exclusivament d'un recull de senyals visuals capturades des de la perspectiva d'una persona, limita l'anàlisi de la informació, eliminant la necessitat d'obtenir informació addicional que pot tenir una greu afectació a la privacitat de la persona.

Aquesta aplicació recull un seguit d'imatges capturades per una càmera portable a baixa freqüència (2fpm), per analitzar-les i determinar la interacció social que hi ha. La interacció social és un factor molt important per a pronosticar la salut física, mental i el benestar de les persones [6]. Per a obtenir patrons de les interaccions socials d'una persona, s'han portat a terme l'observació de la vida d'aquest a llarg termini des d'un punt de vista egocèntric. Es pot concretar que les càmeres d'alta freqüència (per exemple, GoPro o Looxci) no són útils ja que poden registrar imatges només fins 3 hores. En canvi, l'ús de càmeres portables de baixa freqüència (2fpm)(per exemple, Narrative) es poden utilitzar per enregistrar la vida d'una persona, inclòs les seves activitats socials, des d'un punt egocèntric durant un llarg període de temps.

### 2.1.1 Interacció social

Per detectar les interaccions, s'analitzen les imatges per localitzar determinades posicions pròximes a l'usuari, el fet d'evitar oclusions i organitzar orientacions, per a obtenir l'origen de la interacció.

Per a determinar l'existència d'un patró d'interacció, s'utilitza la teoria descrita per Kendon de F-formation [13]. Aquesta teoria defineix un patró que la gent instintivament manté en interactuar amb una altra persona i pot ser mesurat a partir de les distàncies mútues i les orientacions dels subjectes. La F-formation es compon de 3 espais (Fig. 5): p-espai, o-espai i r-espai. El o-espai és un espai vuit convex rodejat per les persones involucrades en la interacció social, on cada participant mira cap al centre de l'espai i no es permet a persones externes en aquest espai. El p-espai és una franja prima que rodeja el o-espai, i conté els participants de la interacció, mentre que el r-espai és l'àrea posterior al p-espai.

En la teoria de F-formation es poden donar diferents configuracions: En el cas de dos participants, les tendències més habituals són vis-avis, en forma de L o un al costat de l'altre. Quan hi ha més de tres participants es crea una formació circular (Fig. 5).

En el cas de les captures egocèntriques, les propietats úniques permeten un enfocament completament nou per a l'anàlisi social.

En referència a la teoria de F-formation, s'analitza la imatge per extreure la seva ubicació i orientació estimades de les cares i aquestes dades s'utilitzen per a calcular la línia de visió per cada cara i així estimar la ubicació 3D d'aquestes (Fig. 4). En obtenir un marc d'estimació de les ubicacions 3D de les persones, s'aplica un algorisme d'agrupació de correlació per fusionar les parelles de persones de grups

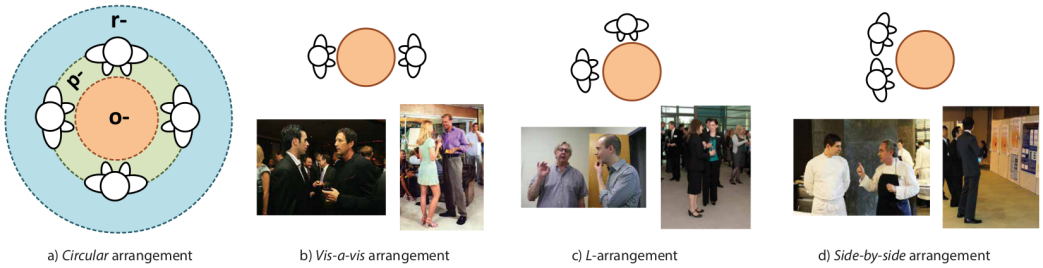


Figura 5: Representació d'espais de F-formation [13]

socials relacionats.

### 2.1.2 Anàlisi d'imatges

Durant el transcurs d'un dia, les persones poden participar en diversos esdeveniments socials. A causa del seu impacte emocional, els esdeveniments socials es podrien considerar com a moments especials que s'haurien de recuperar gràcies a la càmera portable. Però una sola imatge des de la perspectiva de la F-formació aporta una informació limitada, amb la qual es basa l'estat de la interacció social en aquella imatge, on aquest estat té una certa incertesa que fa la decisió poc fiable. Per aquest motiu, es requereix l'anàlisi a escala de seqüència d'imatges per demostrar la participació de les persones en les interaccions socials.

L'aplicació aquí detallada proposa per la detecció i classificació de les interaccions socials dins les seqüències d'imatges egocèntriques l'ús de dos mòduls principals: El primer mòdul té com a objectiu la cerca de les característiques descriptives de la F-formació de cada imatge de la seqüència i prepara les dades a nivell de seqüència. El segon mòdul analitza les característiques resultants del primer modul per a classificar les seqüencies (Fig. 6).

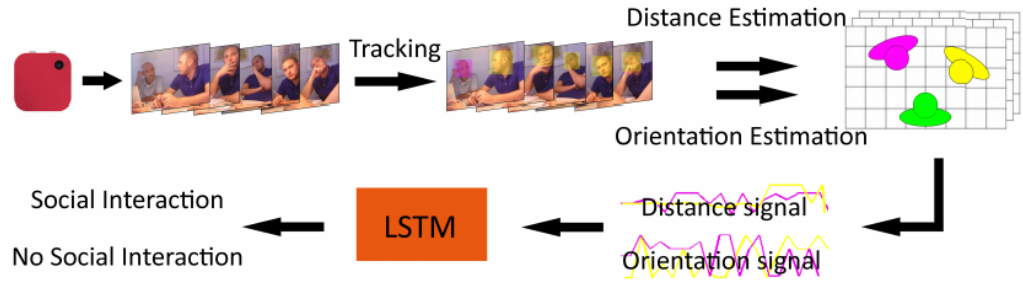


Figura 6: Flux de treball del mètode proposat [7]

#### 2.1.2.1 Extracció de característiques

Per a poder extreure les dades característiques relatives a la F-formació, se segueixen tres mètodes. Primer, es localitza la cara de cada persona que apareix en la imatge.

Després es realitza una estimació de l'orientació de la cara. Per últim, es posa cada persona en escena al llarg de la seqüència d'imatges i es crea una ubicació 3D de cada persona al llarg de la seqüència per tal de construir el conjunt de característiques. Els mètodes es concreten a continuació:

- **Localització de persones dins de la seqüència d'imatges**: El primer pas de tots és detectar i localitzar les persones que estan al voltant de l'usuari de la càmera. Per localitzar al llarg de la seqüència d'imatges les diferents cares, s'utilitza un algorisme de detecció de múltiples cares, el qual ha estat prèviament desenvolupat per seqüències d'imatges egocèntriques [14]. Té com a objectiu calcular la trajectòria de cada persona dins l'escena i mantenir-la a través d'oclusions.

Una vegada realitzada la detecció, se segmenta temporalment la seqüència d'imatges com una porció que estableix prèviament una part on hi ha interacció social i l'altre no. En cada seqüència, per cada cara visible (es determina com llavor), es genera un "tracklet" que es compon d'un conjunt de correspondències al llarg de la seqüència. Posteriorment, s'agrupen els "tracklets" similars en bag-of-tracklets estés (extend) (eBot). Tots els "tracklets" dins un eBot estan destinats a rastrejar a una persona concreta dins la seqüència d'imatges, la qual té la llavor en diferents imatges. Els eBots s'exclouen del conjunt original per mitjà d'una mesura de confiança.

- **Estimació de l'orientació de les cares**: La línia de visió d'una persona es pot calcular aproximadament en estimar la posició del cap de la persona que té davant. Per cada cara detectada, s'amplia el requadre que la delimita, calculat per un petit factor d'ampliació, i es fa una estimació de la postura del cap de l'usuari. El detector es basa en la barreja d'arbres amb un conjunt compartit de parts.

Aquest mètode és capaç de pronosticar l'orientació del cap entre punts de vista discrecionals entre -90º (mira cap a l'esquerra) a 90º (mirar cap a la dreta). Aquest procediment es repeteix per totes les regions de la cara. Per pronosticar la línia de visió se suposa que possiblement es pot mirar des del cantó esquerre fins el cantó dret, tot això dona com a resultat una llibertat de 180º (-90º a 90º). Pel que s'assumeix que les persones que interactuen amb l'usuari de la càmera tenen una posició del cap entre -30º a 30º (Fig. 4).

- **Crear ubicació 3D de les persones**: El model de F-formació es basa en un model de vista d'ocell de l'escena, on cada persona està representada amb dues coordenades (x,z), on x és la posició de la persona en 2D i z la distancia entre ell i la càmera. Per estimar la distància de cada persona, s'entrena un model de regressió que aprèn les relacions de profunditat en una superfície bidimensional. En aquesta aplicació s'assumeix que la distància marge per a una interacció social són 150cm.

#### 2.1.2.2 Classificació d'interacció social amb LSTM

Les característiques descrites en l'apartat anterior codifiquen una instantània local en el temps. Però és important l'anàlisi en el canvi temporal. Es modela el problema utilitzant un classificador de Xarxes Neuronals Recurrent (RNN) particulars, anomenats LSTM, per aprofitar la seva capacitat de l'evolució temporal dels des-

criptors. S'aplicaran les característiques per a la classificació de les seqüències en interacció social.

Per a una classificació binària, de seqüències d'imatges egocèntriques, es proposa entrenar una xarxa LSTM introduint-li els vectors característics de cada seqüència d'imatges amb la informació de la distància com la de l'orientació. El sistema tindrà que aprendré a classificar seqüències de diferents longituds per interactuar o no per mitjà de l'anàlisi dels dos vectors de característiques associats a cada seqüència. Per aquest motiu el sistema necessita aprendre a protegir els continguts de les cel·les de memòria inclús contra deriva d'estat intern menor.

La capa oculta conté diverses cel·les de memòria totalment interconnectades i totalment connectades a la resta de la xarxa. Les portes d'entrada i sortida s'utilitzen entrades d'altres cel·les de memòria per decidir si accedir a informació concreta dins una cel·la de memòria. Sent la i-ésima cel·la de memòria $c_i$, en temps $t$, la sortida de $c_i$ i $y_i^c(t)$ es calcula com:

$$y^{c_i}(t) = y^{out_i}(t)h(s_{c_i}(t)),$$

on l'estat intern $s_{c_i}(t)$ és:

$$s_{c_i}(0) = 0$$
$$s_{c_i}(t) = s_{c_i}(t-1) + y^{in_i}(t)g(net_{c_i}(t)) \;\; for \;\; t > 0,$$

on $in_i$ i $out_i$, són la porta d'entrada i la porta de sortida de la cel·la, respectivament. $g$ és una funció diferenciable que anul·la $net_{c_i}$, i $h$ és una funció diferenciable que escala la sortida de la cel·la de memòria calculada a partir de l'estat intern $s_{c_i}$.

Cada unitat de la capa d'entrada rep la distància i orientació d'una persona per l'usuari de la càmera en un marc determinat. La capacitació de la xarxa LSTM és essencial per una classificació precisa de les entrades seqüencials.

### 2.1.3 Aplicacions

Es proposa una canalització completa per detectar interaccions socials en seqüències d'imatges egocèntriques capturades per mitjà d'una càmera portable de baixa freqüència (2fpm). Els resultats experimentals han demostrat que el mètode proposat ha obtingut una alta precisió en el propòsit considerat (78% utilitzant SGD).

Aquest treball té potencial d'aplicacions importants en els camps de la medicina preventiva i la interacció humana-computació com, per exemple, l'entrenament de memòria amb persones afectades per deteriorament cognitiu lleu i la teràpia robòtica per a nens afectats per autisme. Informació de les persones amb les que l'usuari de la càmera interactua, seria d'importància per a la creació de *serious games*, com els proposats a [2].

## 2.2 Caracterització d'estil social

Aquesta aplicació proposa un sistema per detectar de manera automàtica patrons socials per mitjà de les característiques de les imatges capturades des de les càmeres

portables dels usuaris. Per portar a terme aquest propòsit, es plantegen tres passos principals els quals són:

-**Detectar les persones** amb les que interactua l'usuari amb la càmera portable.

-**Categoritzar la interacció social** obtingudes del pas anterior per detectar si la interacció és formal o informal. Amb els dos passos esmentats, es fa una anàlisi a escala d'esdeveniment individual per crear una sèrie temporal multi-dimensional de característiques rellevants.

Cada dimensió correspon a un conjunt de característiques rellevants per cada tasca i s'utilitza una xarxa neuronal LSTM per a la classificació de les sèries temporals.

-**Analitzar les interaccions per cada persona** on es recullen les recurrències de cada persona amb qui interactua l'usuari, per a fer una comprensió total de la diversitat i freqüència de les relacions socials. Per a posar a prova aquesta aplicació, s'ha fet ús de les imatges realitzades durant un mes d'un usuari.

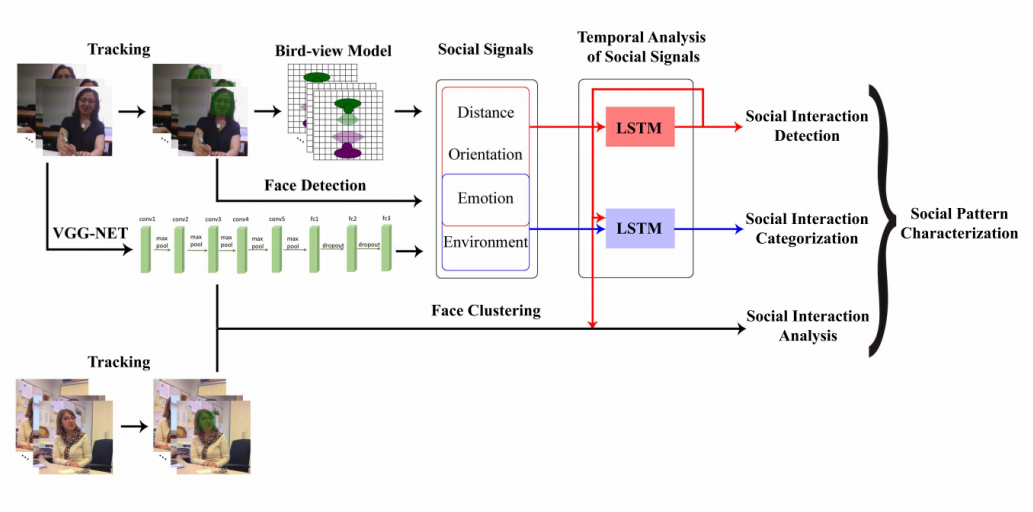


Figura 7: Flux de treball proposat [8]

El patró social de l'usuari es genera pel resultat del descobriment de les persones recurrents, en el conjunt de dades i en la quantificació de la freqüència, la diversitat i el tipus d'interacció socials succeïdes amb diferents individus.

### 2.2.1 Detecció d'interacció social

Utilitzant la metodologia detallada en [7] i comentada en l'aplicació anterior, primer se segmenta el flux de les imatges en esdeveniments individuals i se seleccionen els possibles esdeveniments socials. En cada esdeveniment social, es rastregen les cares per mitjà d'un algorisme de seguiment de múltiples cares [14].

Per detectar la interacció social per cada persona rastrejada s'utilitza una classificació binaria de les sèries temporals (interacció vs. no interacció), on la dimensió de la sèrie temporal correspon al nombre de senyals socials seleccionades per descriure

una interacció social.

Els valors obtinguts en la detecció social són la distància ($\varphi d$) respecte a la càmera portable, l'orientació de cara detectada segons la seva rotació (yaw)($\varphi z$), Fig. 8, igual que es proposa en [7]. Però aquesta aplicació explota més punts per determinar l'impacte de l'orientació de la cara segons els termes d'inclinació (pitch)($\varphi y$), Fig. 8, i balanceig (roll)($\varphi x$), Fig. 8, així com d'expressió facial ($\varphi e$).

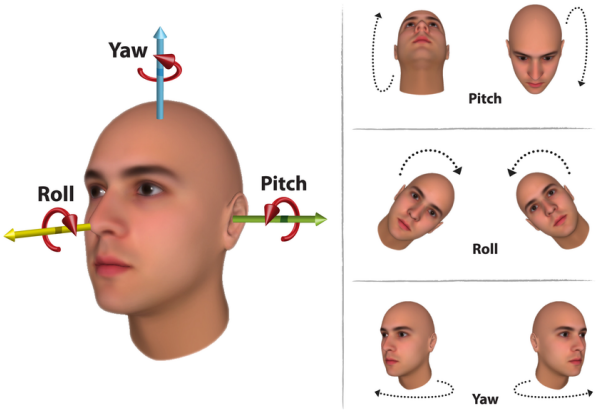


Figura 8: Determinació de coordenades [7]

Les expressions facials es representen com un vector de probabilitats per cada una de les 8 expressions facials diferents associades a les emocions en la cultura occidental.

Per una persona donada $p_i$, el índex d'expressió facial dominant seria:

$$\varphi_c = arg \max_{k\epsilon 1,\ldots,8} e_k(p_i) \tag{2.1}$$

Es considera el valor de l'expressió facial. El conjunt complet de característiques és una sèrie temporal de 5 dimensions (distancia, roll, pitch, yaw i expressió facial) que representen l'evolució temporal de les característiques d'interacció j-éssima al llarg dels temps, extretes per separat per cada cara rastrejada:

$$\varphi^{\tau}_{detection} = (\varphi^{\tau}_d, \varphi^{\tau}_z, \varphi^{\tau}_y, \varphi^{\tau}_x, \varphi^{\tau}_c), \tau = 1, 2, ....$$

### 2.2.2 Categorització de interacció social

La definició sociològica de reunions formals i informals, com dos grans categories d'interacció social des de la perspectiva de la visió per computació, suggereix que les característiques ambientals i l'expressió facial mostren un poder discriminatori en la categorització.

- **Característiques mediambientals**: Cada component del vector de característiques extret d'una xarxa neuronal convolucional (CNN) té algun contingut

semàntic, que pot considerar-se com un bon representant de l'entorn en una imatge. Per reduir l'excés de dimensions del vector de característiques obtingut de la xarxa CNN (4096D), s'aplica un enfocament per reescriure el vector envers paraules discretes [16].

Posteriorment, s'aplica anàlisi de components principals (PCA) per a mantenir el 95% de la informació més important de la matriu resultant, on tot això condueix a un vector de característiques de 35 dimensions com $\varphi_g \epsilon R^{35}$.

- **Expressió facial**: Les característiques de l'expressió facial en aquesta tasca s'extreuen com la mitja de les expressions facials del nombre total de persones detectades en la imatge de la seqüència:

$$\varphi_{c,ind} = \frac{1}{n}\sum_{i=1}^{n} e_{ind}(p_i), ind = 1, ..., 8. \tag{2.2}$$

Aquest punt de vista té en compte l'evolució temporal de les característiques d'expressió facial i ambiental en ser modelades com a sèries temporal multi-dimensional:

$$\varphi^{\tau}_{categoritzacion} \epsilon R^{43} = (\varphi^{\tau}_{g}, \varphi^{\tau}_{e}), \tau = 1, 2, ...$$

on es basa en el LSTM per a la classificació binària de cada sèrie de temps en una reunió formal o informal.

### 2.2.3 Caracterització de la interacció

Per poder caracteritzar el patró social d'una persona s'ha de portar a terme l'anàlisi de les interaccions socials per mitjà de diversos esdeveniments socials durant un llarg període de temps. Això implica la necessitat de definir tres conceptes per tal de poder concretar la naturalesa de les interaccions socials. Aquests conceptes són: la freqüència, la diversitat i la duració.

- **Freqüència**: Es defineix com a la taxa de interaccions formals (informals), on $I$ és una persona normalitzada pel número total de interaccions.

$$F_{f(inf)} = \#I_{f(inf)}/\#dies$$

- **Diversitat**: Comprova la diferència de les interaccions socials d'una persona. Aquest concepte es defineix com l'exponencial de l'entropia de Shannon calculada amb logaritmes naturals:

$$D = 1/2exp(-\sum_{i\epsilon\{f,inf\}} A_i ln(A_i)),$$

on $A$ indica si la majoria de les interaccions socials d'una persona són formals (informals), respectivament: $A_{f(inf)} = \#I_{f(inf)}/\#I$).

Tenint en compte que si una persona té el mateix nombre d'interaccions formals i informals ($A_{formal} = A_{informal} = 0,5$), llavors $D = 1$.

- **Duració**: És la longitud d'una interacció social. Es defineix com $L(i)$ per cada interacció social i de l'usuari. És proporcional a la longitud de la seqüència corresponent a la interacció social.

Per exemple: $L(i) = \tau(i)r$, on $\tau(i)$ és el nombre d'imatges de la interacció i-ésima i $r$ és la velocitat d'imatges de la càmera.

### 2.2.4 Aplicació de l'algorisme de caracterització de la interacció

Es proposa una canalització completa per la caracterització del patró social d'un usuari que utilitza una càmera portable durant un llarg període de temps.

S'ha comprovat aquesta aplicació realitzant dos experiments on s'han analitzat les imatges en diferents formats, per tal de comprovar quin és el més eficient al detectar interaccions socials i categoritzar interaccions socials.

S'han creat les següents configuracions per comprovar la detecció d'interaccions socials:

SID1: Distance + Yaw

SID2: Distance + Yaw + Pitch + Roll

SID3: Distance + Yaw + Facial expression

SID4: Distance + Yaw + Pitch + Roll + Facial expressions

S'han creat les següents configuracions per comprovar la categorització de les interaccions socials:

SIC1: Environmental (VGG)

SIC2: Environmental (VGG-finetuned)

SIC3: Environmental (VGG-finetuned) + Facial expressions

Els resultats d'aquests dos experiments han estat prometedors, on les configuracions **SID4** i **SIC3** han estat les millors amb una 'precisió de 82,55% i 91,15%, respectivament. L'anàlisi a nivell de seqüència que utilitza la LSTM aporta un major rendiment en ambdues tasques.

Les possibles aplicacions en el camp de la medicina preventiva són molt importants, ja que per exemple es poden estudiar els patrons socials dels pacients afectats per depressió, també de les persones grans i de les persones afectades per algun tipus de trauma.

## 2.3 Persones del voltant

Donat un flux d'imatges sense restriccions capturades per una càmera fotogràfica usable (2fpm), es proposa un enfocament no supervisat de baix a dalt per a agrupacions automàtiques que apareixen cares en les identitats individuals presents en

aquestes dades. El problema és difícil ja que les imatges s'adquireixen en condicions del món real; d'aquí l'aspecte visible de les persones en les imatges experimenta variacions intensives. El nostre projecte de canalització consisteix en organitzar primer la foto-stream en esdeveniments, més tard, localitzar l'aparició de diverses persones en ells, i finalment, agrupar diversos aspectes de la mateixa persona en diferents esdeveniments. Els resultats experimentals realitzats en un conjunt de dades adquirits mitjançant una càmera fotogràfica durant un mes, demostren l'efectivitat si es proposa l'enfocament per al propòsit considerat.

### 2.3.1 Treball realitzat

El clúster de rostres és un problema poc complex i una gran quantitat de treball en la literatura s'ha centrat a trobar com explotar les característiques del conjunt de dades o de l'aplicació particular per restringir-la. Les aplicacions més freqüents són l'etiquetatge interactiu d'àlbums de fotos i organitzacions de vídeo. En el context de la cara descobriment en àlbums de fotos, Lee et al. va introduir una nova restricció coneguda com a context social de les persones que van passar a ser conegudes, de manera que sovint apareixen persones del mateix context social. Per exemple, les cares dels membres de la família solen coincidir, fins i tot, en diferents fotografies. El sistema fa un primer detector per a cada individu i, posteriorment, utilitza el detector per descobrir clústers de cara nous aprofitant les restriccions de coincidència. En el mateix escenari, Zhu et al. va presentar una distància de rang-ordre per mesurar la dissimilaritat entre dues cares. Aquesta obra explora el fet que les cares de la mateixa persona solen formar subclústers propers a l'espai de funcions. Una idea semblant és proposada per Xia et al., que va aprofitar dues restriccions: només un individu pot aparèixer una vegada a la imatge, i la quantitat d'instàncies d'una mateixa persona ha de ser inferior a la quantitat total d'imatges. El problema es forma llavors com un K-Means restringit, que es resol a través de l'estratègia d'optimització de xarxes lineal de flux mínim de costos. La imposició de restriccions per aconseguir un clúster més precís s'observa en diversos altres treballs que intenten agrupar cares en vídeos. Xiao et al. va proposar una Representació de rang baix de blocs ponderats (WBSLRR) que aprèn una representació de dades de baix rang, tot considerant dues restriccions anteriors definides. En primer lloc, la restricció de la pista interna estableix que les dues cares de la mateixa carícia pertanyen a la mateixa persona. Per tant, el clúster es realitza per primera vegada en rutes de cara en comptes de cares individuals. En segon lloc, la restricció entre traces que estableix rostres de rostres que pertanyen a cares que apareixen en el mateix marc, no pertany a la mateixa persona. Una idea semblant ha estat emprada per Cinbis et al. , per obtenir una mètrica de distància per a la identificació de cara en vídeos que junten cares en una relació de trama interna, i allunyen els que es troben en relació intertrau. Més recentment, com en moltes altres tasques de visió per computadora, les característiques profundes van demostrar la seva eficàcia en la representació de dades per agrupacions facials. No obstant això, es supervisen els enfocaments basats en l'aprenentatge profund i, per tant, requereixen una etapa d'aprenentatge prèvia amb cares etiquetades amb identitat. Per tant, són més adequats per a la reidentificació facial.

# 3 Re-identificació de Persones

En aquest apartat és descriu el sistema proposat per a la re-identificació de persones. Aquest sistema analitza diverses imatges amb l'objectiu de localitzar rostres de persones i classificar-les per mitjà de característiques facials i/o contextuals. Per a arribar a l'objectiu esmentat aquest apartat es divideix en diversos apartats, dels quals es farà una breu explicació.

En primer lugar, per a la **detecció** de cares proposam explicar l'algorisme Viola-Jones [18]. Aquest localitza els rostres de les persones en les imatges. Amb els rostres localitzats, es farà l'**extracció de característiques facials** d'aquests, amb les quals es crearà un vector descriptiu per cada rostre. Posteriorment es localitza la zona del pit de la persona, per tal de caracteritzar la roba. En generar els dos vectors característics, rostre i roba, es realitzaran **agrupaments** d'aquests vectors per a poder classificar els diferents rostres, generant clústers de rostres iguals.

En aquest treball proposem el següent model: la classificació de les imatges es sobre les característiques facials i contextuals - roba. Les cares es descriuen amb el vector obtingut amb OpenFace [20], la roba amb descriptors de color HSV. Com a mètode d'agrupament es proposa l'agrupament Jeràrquic [21][25], amb el coeficient Pearson [23] per a comprovar les similituds dels rostres.

A continuació, es fa una breu explicació de les tecnologies empleades.

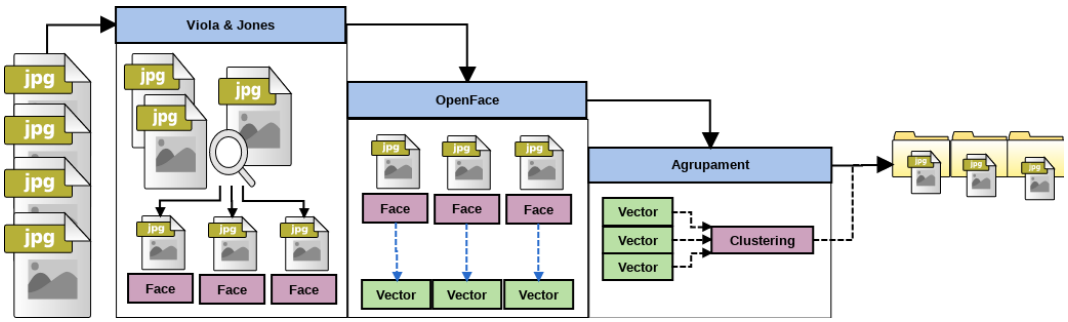


Figura 9: Esquema bàsic de la metodologia

La Fig.9 mostra el recorregut de les imatges des que entren al sistema fins que surten organitzades per carpetes segons les cares. Aquest esquema és una representació bàsica del tractament de la informació, on en cada punt s'apliquen altres mecanismes per augmentar la precisió del sistema.

En la Fig. 29 mostra el recorregut de la metodologia amb més detall i amb un resultat final de l'organització de les imatges. Aquesta figura està ubicada al final d'aquest apartat.

## 3.1 Detecció de cares

Per tal de localitzar cares ubicades a diverses fotografies, hi ha l'algorisme de Viola & Jones, el qual consta de dues parts principals: un classificador en cascada, que

facilita una discriminació ràpida i un entrenador de classificadors basat en Adaboost. Aquest detector de cares és molt utilitzat pel seu baix cost computacional i una alta probabilitat d'obtenir veritables positius.

El classificador en cascada és un conjunt de filtres ordenats de menys cost a més cost computacional. D'aquesta manera es pot fer un descart ràpid de les imatges de manera eficient. Per extreure les característiques de la imatge, s'utilitzen les "Haar features", que es mostren a la Fig. 10. Aquest és un nucli convergent, on cada característica és un valor únic obtingut de restar la suma de píxels sota el rectangle blanc de la suma de píxels en el rectangle negre.

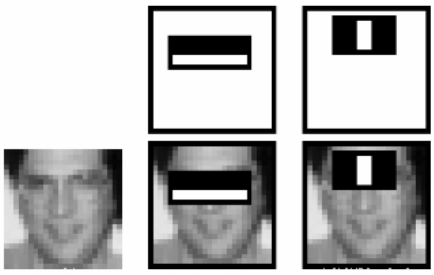

Figura 10: Haar features

L'algorisme extreu les característiques de la imatge en escala de grisos per mitjà dels "Haar features"abans esmentats i gràcies als diferents classificadors en cascada va descartant les parts de la imatge que no tenen cares fins a arribar a l'últim filtre, on llavors, les zones que no han estat descartades es considera que representen cares.

El codi implementat per l'excussió de l'algoritme de Viola & Jones, és el que és mostra en la Fig. 11, a continuació:

```
faces = []
for cascade in face_cascades:
    f = cascade.detectMultiScale(gray, scaleFactor = scale_factor,
                                 minNeighbors = min_neighbors,
                                 flags = flags,
                                 minSize = (size_min,size_min),
                                 maxSize = (size_max,size_max))
    f = list(f)
    if f: faces += f

centros = []
for i, ( x, y, w, h ) in enumerate(faces):
```

Figura 11: Codi de l'algorisme Viola & Jones

Com es pot observar a la Fig. 11, la funció per utilitzar el mètode de Viola & Jones és la "detecMultiScale()", la qual es crida des de un objecte creat amb "cv2.CascadeClassifier('arxiu.xml')". En aquest cas, aquest objecte és "cascade" i l'arxiu xml és on es guarden els valors entrenats dels diferents filtres en cascada. Per aquest motiu, aquest algorisme es podria utilitzar per a la detecció d'altres elements i no sols cares, ja que es podria entrenar sobre diferent objectes.

A part de la selecció de l'arxiu xml, l'algorisme disposa d'altres paràmetres per concretar més la cerca, els quals es mostren en la Fig. 11, com a scaleFactor,

```
24  # Extreure cares de les imatges
25  def imageFaces(image, outputDir,  mutex = None,
26                 face_cascade_paths=['haarcascades/haarcascade_frontalface_alt2.xml', 'haarcascades/haarcascade_profileface.xml'],
27                 scale_factor=1.1, min_neighbors=13,
28                 exts=['*.jpg','*.jpeg','*.png'],
29                 min_size=-1, max_size=-1, flags = cv2.CASCADE_SCALE_IMAGE, width_view = 0.9, marc=False):
30
```

```
# Possar diferents filtres pel classificador
face_cascades = [cv2.CascadeClassifier(os.path.expanduser(path)) for path in face_cascade_paths]
```

Figura 12: Combinació de filtres *.xml

minNeighbors, flags, minSize i maxSize, els quals s'explicaran en més detall en la part d'experiments, ja que tenen un valor important per la velocitat d'execució.

Concretament, en aquest projecte s'han utilitzat dos arxius ja entrenats de la llibreria OpenCV els quals són "haarcascade_frontalface_alt2.xmlï "haarcascade_frontalcatface.xml", on el primer s'utilitza per detectar cares de manera frontal i el segon s'utilitza per detectar cares de perfil. En la Fig. 12, es pot observar com es combinen els diferents filtres dins la variable "face_cascades", la qual posteriorment serà utilitzada per buscar cares (Fig. 11). OpenCV també disposa d'altres arxius xml per la detecció d'ulls, boca i objectes, pel que és una llibreria lliure molt completa.

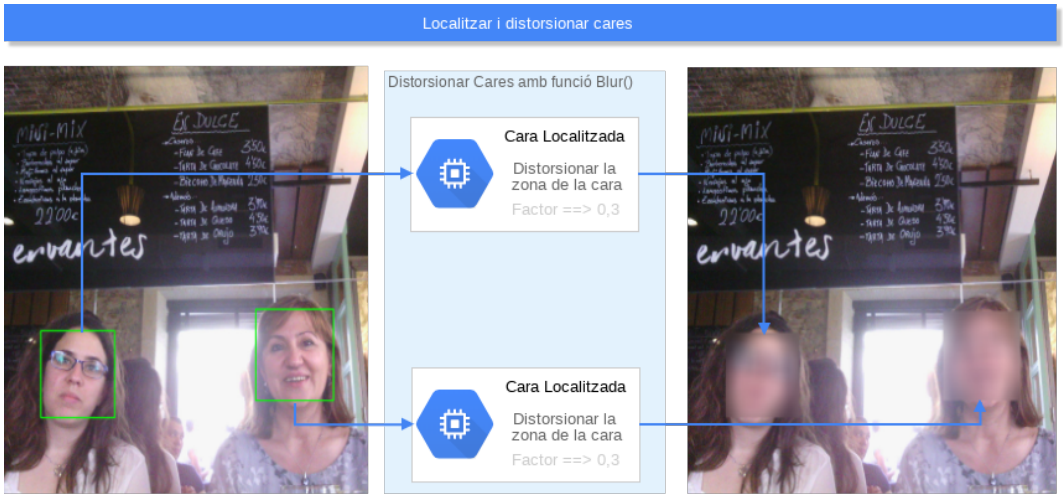


Figura 13: Localització i distorsió de cares

En la Fig. 13 es pot observar com es localitzen les zones de la imatge que contenen un rostre i posteriorment es distorsionen aquestes zones per tal de privatitzar la imatge. Com es pot observar en la figura per tal de distorsionar les zones s'aplica una funció anomenada "blur()", la qual està incorporada a les llibreries OpenCv [17] i estarà explicada en el pròxim punt.

**Paràmetres de l'algorisme de Viola & Jones**

Per tal de concretar el funcionament d'aquest algorisme a continuació es detallen els paràmetres de Viola & Jones, el qual està incorporat a la llibreria OpenCV per mitjà de la funció *cv2.CascadeClassifier.detectMultiScale(image[, scaleFactor[, minNeighbors[, flags[, minSize[, maxSize]]]]])* .

- **image**: Paràmetre obligatori, ja que és la representació de la imatge en matriu en escala de grisos.

- **scaleFactor**: Paràmetre opcional, especifica en quina proporció es redueixen les dimensions de la imatge segons el factor d'escala que es doni. A major valor augmenta la velocitat d'execució, però pot minorar la precisió de la detecció. Al contrari, un factor d'escala més petit augmenta el temps d'execució, però pot produir una major detecció de falsos positius.

- **minNeighbors**: Paràmetre opcional, controla el nombre mínim de píxels delimitadors detectats en una determinada regió. A major valor dona com a resultat la detecció de cares de major qualitat, però pot reduir la detecció d'altres cares.

- **flags**: Paràmetre opcional, per indicar un tipus concret de *haar feature*, és més utilitzat en versions anteriors, actualment s'ha indicar *cv2.CASCADE_SCALE_IMAGE*.

- **minSize**: Paràmetre opcional, aquest valor garanteix que el requadre delimitador detectat tingui una **dimensió mínima** d'ample $x$ alt de píxels. Per exemple (10, 10) dona com a resultat que la detecció mínima seran requadres de $10x10$ píxels.

- **maxSize**: Paràmetre opcional, aquest valor garanteix que el requadre delimitador detectat tingui una **dimensió màxima** d'ample $x$ alt de píxels. Per exemple (100,100) dona com a resultat que la detecció màxima seran requadres de $100x100$ píxels.

Al definir els valors minSize i maxSize, es pot arribar a millorar de manera molt dràstica la velocitat d'execució, ja que es defineixen les dimensions de cerca, com és petit sigui el valor en píxels entre minSize i maxSize, més velocitat d'execució tindrà, però s'ha de tenir present que les cares que no estiguin entre aquests valors seran descartades (No es detectaran).

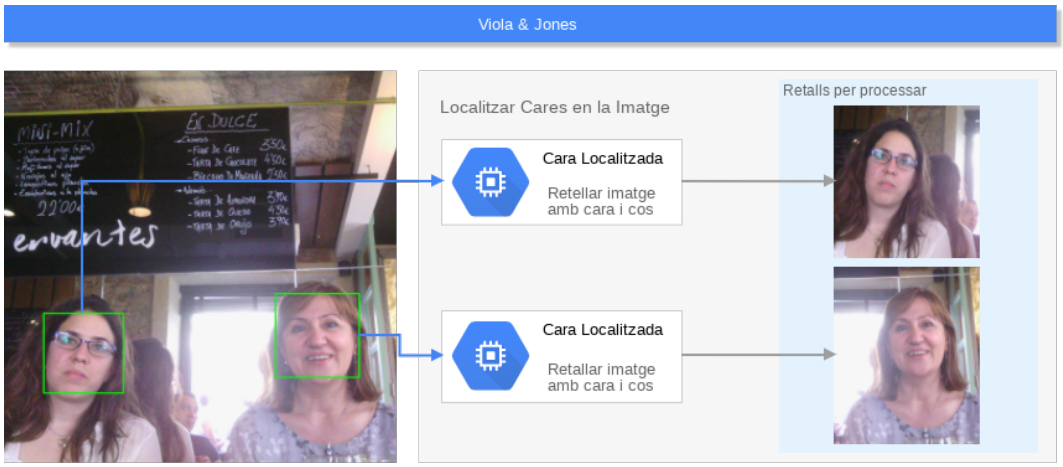


Figura 14: Localització de cares amb Viola & Jones. Es mostra com es localitzen les cares en una imatge i per mitjà de les proporcions del valor *Phi*.

### 3.1.1 Aplicació: Privatització de cares

La incorporació de les càmeres portables en el món de la investigació ha aportat una perspectiva innovadora amb la qual s'extreu informació molt valuosa. Per contrapartida, s'ha de tenir present que aquesta informació té una gran sensibilitat en tractar imatges de caràcter personal. La legislació actual, regulada per la Llei Orgànica 15/1999 del 13 de Desembre envers la protecció de dades de caràcter personal, concreta la necessitat inherent de tenir el consentiment de la persona que apareix en la imatge, tal com cita l'article 6.1 de l'esmentada llei, on aquest diu textualment: *"El tratamiento de los datos de carácter personal requerirá el consentimiento inequívoco del afectado, salvo que la ley disponga otra cosa."*.

Pel motiu abans exposat, s'han creat eines per a distorsionar les cares de les persones que apareixen en les imatges. Aquestes eines tenen com a finalitat ocultar el rostre de la persona fotografiada, perquè aquest continuí amb el seu anonimat.

Per portar a terme la privatització de les imatges, s'ha seguit el següent procediment: Primer s'han localitzat les cares de les imatges per mitjà de l'algorisme de Viola & Jones [18] integrat a les llibreries d'OpenCV [17]. I posteriorment, s'ha aplicat una funció de distorsió (Blur) per tal de privatitzar la zona on s'ha localitzat la imatge.

**Procediment:**

Una vegada s'ha localitzat, si hi ha, la cara o cares dins la imatge, en aquesta s'aplica una funció per a distorsionar la zona on es troba la cara. Com es pot observar en la Fig. 13, hi ha una imatge a l'esquerra amb dues cares, les quals són detectades per mitjà de Viola & Jones. Es pot observar com les zones emmarcades són processades i com el resultat apareix en la imatge de la dreta, en la qual les zones localitzades amb la cara apareixen difuminades de tal manera que no es pot reconèixer la cara.

Per a portar a terme la distorsió de les cares localitzades, s'utilitza la funció "cv2.blur(src, ksize[, dst[, anchor[, borderType]]])", Fig. 15, la qual està implementada en les llibreries de OpenCV [17]. Aquesta funció té diversos paràmetres, però els paràmetres "**src**ï "**ksize**"són més que suficients per a complir amb la finalitat desitjada.

Bàsicament el paràmetre de "**src**" és la imatge que es vol distorsionar, la qual està representada en forma de matriu. La funció "blur" distorsiona tota la imatge d'entrada. Per aquest motiu, el primer que es fa és localitzar la cara dins la imatge, retallar-la per introduir-la en la funció "blur" i una vegada distorsionat el retall col·locar aquest en la imatge original (mirar Fig. 13).

L'altre paràmetre "**ksize**" és per determinar la dimensió del nucli de distorsió, per exemple (33,33). Per tal de facilitar la introducció d'aquest valor, es crea una fórmula, on es multiplica l'altura de la zona de la cara per un factor (línia 268 de la Fig. 15). El factor que s'utilitza és un valor entre el 0 i el 1. Aixi és pot simplificar l'entrada d'aquest valor i determinar els nivells de distorsió.

Com es pot observar en la Fig. 16 un factor òptim perquè no es reconegui a la

```
        #Equalizar Histograma per millorar qualitat per lluminositat
        cv2.equalizeHist(gray, gray);

        faces = []
        for cascade in face_cascades:
            f = cascade.detectMultiScale(gray, scaleFactor = scale_factor,
                                         minNeighbors = min_neighbors, flags = flags, minSize = (size_min,size_min), maxSize = (size_max,size_max))
            f = list(f)
            if f: faces += f

        centros = []
        nfaces = 0
        for i, ( x, y, w, h ) in enumerate(faces):
            unico =True # Controla que no haya recuedros repetidos

            #Para Gesionar los recuadros repetidos pulir
            for cx,cy in centros:
                if cx > x and cy > y and cx < (x+w) and cy < (y+h):
                        unico = False # Recuadro repetido

            if unico:
                nfaces += 1
                # Guardar recorte expandido sin recuadro
                img_crop_ext = img[y:y+h,x:x+w]

                # Distorsionar cara se multiplica longuitud cuadrado cara por nivel
                img_blur = cv2.blur(img,(int(h*blur_lvl),int(h*blur_lvl)))

                # Possar la zona amb la cara distorsionada
                img[y:y+h,x:x+w] = img_blur[y:y+h,x:x+w]

        #outfname = outputDir+"/"+image.split('/')[-1]
        outfname = outputDir+'/'+img_basename+'-'+str(nfaces)+'.'+img_ext
        # Bloquejar el thread per poder guardar la foto
        if mutex is not None:
            mutex.acquire()
        cv2.imwrite(outfname, img)

        # Desbloquejar tots els threads
        if mutex is not None:
            mutex.release()
```

Figura 15: Codi per localitzar i distorsionar cares

Figura 16: Distorsió cares : Original, factor 0.1, factor 0.2, factor 0.3 i factor 0.6

persona seria el 0.3, però com és clar, aquest factor també ha de ser el valor òptim segons les necessitats del problema.

Una vegada detectades les cares a la col·lecció d'imatges per una persona, volem identificar-les per recuperar la mateixa persona en diferents esdeveniments. L'objectiu és crear un *àlbum* d'interaccions socials amb la forma de *llista de contactes*. Per poder crear un àlbum amb les persones ben catalogades, s'han de classificar les imatges obtingudes del portador de la càmera portable. D'aquest conjunt d'imatges, se seleccionen sols les imatges on apareixen rostres de persones, de les quals s'extreuen els rostres amb part del seu cos.

Amb els rostres obtinguts, s'obtenen les seves referències per a poder classificar les cares per mitjà de diferents sistemes d'agrupació. També, es preten comprovar si utilitzant els valors del context, extrets amb els rostres, els rostres es poden classificar de forma més eficient al combinar les característiques dels rostres amb les característiques del context.

## 3.2 Extracció de característiques

En aquest apartat explicarà com s'extreuen els valors característics de les zones localitzades de rostre i roba. També explicarem les eines utilitzades per generar els vectors descriptius per poder representar les característiques i poder processar les imatges.

### 3.2.1 Característiques facials

OpenFace és un conjunt de llibreries lliures ubicades a GitHub, concretament a la web: "http://cmusatyalab.github.io/openface/ ", que s'utilitzen per a obtenir les característiques concretes de les cares per mitjà de Deep Learning, utilitzant una xarxa ja entrenada.

Existeixen diverses llibreries destinades a obtenir les característiques d'una cara. Però, moltes d'aquestes són llibreries d'ús privat, ja que han estat explotades per un ús comercial. En canvi, OpenFace és una llibreria lliure que s'ha anat ampliant. El seu codi està directament penjat al sistema GitHub, i es pot utilitzar segons les necessitats de cada problema.

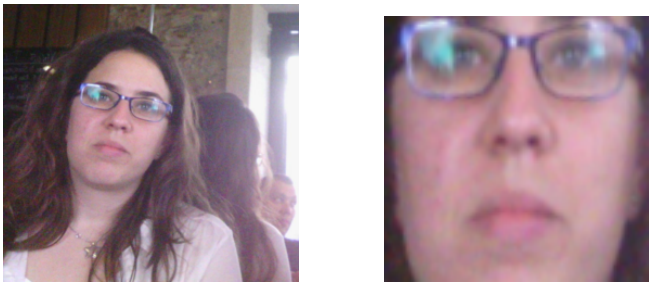

Figura 17: Imatge original i cara alineada

Per realitzar l'obtenció de les característiques d'una cara per mitjà de Deep Learning, és semblant a l'algorisme de Viola & Jones, ja que el sistema s'entrena prèviament. OpenFace ja consta d'un entrenament i prova previs de més de 6.000 imatges amb una precisió del 87%. La representació de cada cara consta d'un vector amb 128 valors compresos entre -0.4 a 0.4 , pel que cada cara es representa amb 128 bytes. Aquest vector és el que s'utilitza com a vector característic. Posteriorment s'utilitza per a la classificació de les cares per mitjà d'agrupament.

A continuació, en la Fig. 19, és mostra la cara localitzada i retallada d'una fotografia més general per mitjà de Viola & Jones. La cara mostrada sol és un retall de la imatge localitzada, però per poder identificar l'esmentada cara s'ha d'obtenir un vector característic. Gràcies a les funcions d'OpenFaces, s'obté el vector característic de 128 bytes que es mostra en la Fig. 20.

Per poder portar a terme l'obtenció del vector característic, és imprescindible alinear la cara, Fig. 17 i 19, per a reduir tot el possible l'obtenció de valors erronis per culpa de la posició de la cara.

```
50      if aviso:
51          print("Processing {}.".format(imgPath))
52
53      bgrImg = cv2.imread(imgPath)
54
55      if bgrImg is None:
56          #raise Exception("Unable to load image: {}".format(imgPath))
57          #return ['Error'] # En caso de errores
58          return None #'Error'
59
60      rgbImg = cv2.cvtColor(bgrImg, cv2.COLOR_BGR2RGB)
61
62      if aviso:
63          print("  + Original size: {}".format(rgbImg.shape))
64
65      start = time.time()
66      bb = align.getLargestFaceBoundingBox(rgbImg)
67      if bb is None:
68          #raise Exception("Unable to find a face: {}".format(imgPath))
69          #return ['Error'] # En caso de error
70          return None #'Error'
71
72      if aviso:
73          print("  + Face detection took {} seconds.".format(time.time() - start))
74
75      start = time.time()
76      alignedFace = align.align(imgDim, rgbImg, bb,
77                                landmarkIndices=openface.AlignDlib.OUTER_EYES_AND_NOSE)
78      if alignedFace is None:
79          #raise Exception("Unable to align image: {}".format(imgPath))
80          #return ['Error'] #En caso de error
81          return None #'Error'
82
83      if aviso:
84          print("  + Face alignment took {} seconds.".format(time.time() - start))
85
86      start = time.time()
87      rep = net.forward(alignedFace)
88
89      if aviso:
90          print("  + OpenFace forward pass took {} seconds.".format(time.time() - start))
91          #print("Representation:")
92          #print(rep)
93          print("-----\n")
94
```

Figura 18: Codi OpenFace per extreure vector característic

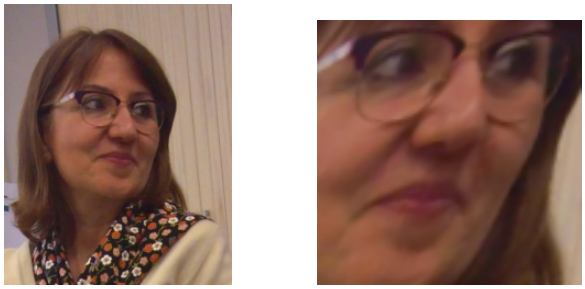

Figura 19: Imatge de mostra Original i Alineada

Per alinear la cara, OpenFace disposa d'eines per aquest fet. En la Fig. 18 es pot observar el codi que s'utilitza per obtenir el vector característic d'una cara. La imatge que entra no ha d'estar alineada, ja que en la línia 76 s'alinea la cara detectada i es comencen a extreure els valors del vector característic.

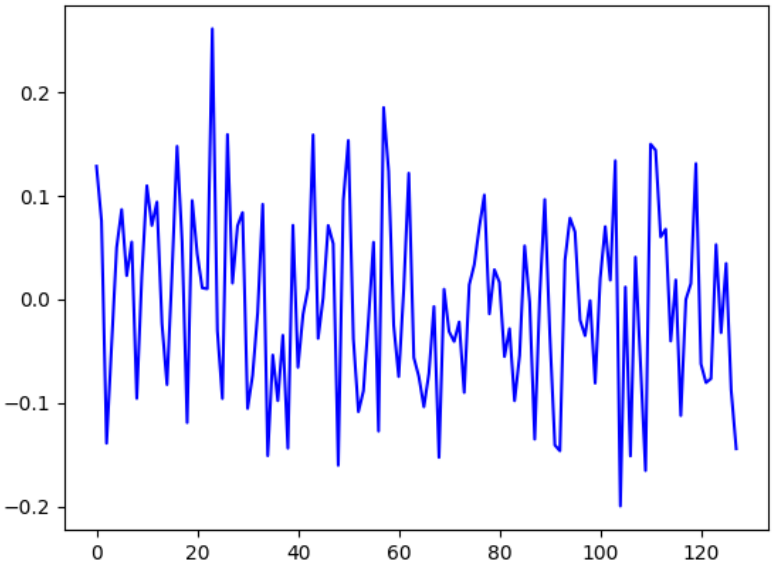


Figura 20: Vector característic de la Figura 19

Amb la Fig. 20 es poden observar les característiques concretes de la cara mostrada a la Fig. 19 junt amb la seva alineació. En obtenir el vector característic de cada cara, aquest per mitjà de tècniques d'agrupament, s'utilitza per a la identificació de les persones. Per realitzar l'agrupament amb aquest tipus de vector, no és necessari fer una normalització, ja que tots els vectors extrets per mitjà d'OpenFace ja estan normalitzats entre els valors 0.4 a -0.4.

### 3.2.2 Característiques contextuals - Discriminador de Roba

Per a comprovar si les persones porten la mateixa roba en les diferents fotografies, s'ha utilitzat un vector característic, el qual extreu els valors dels histogrames, concretament dels píxels corresponents a la zona de la roba.

La zona que es determina com a roba, s'obté gràcies a les coordenades de la cara obtingudes per l'algoritme de Viola & Jones. Per tal de detectar l'àrea del tors de la persona d'on extraiem els descriptors de roba, ens basam en el Golden ràtio *Phi*. Aquest es també conegut com a número Aureo ($\varphi$) es tracta d'un número algebraic irracional que conte moltes propietats, com per exemple la relació o proporció entre dos segments d'una recta, és a dir, una construcció geomètrica. Aquesta proporció es troba en diverses figures geomètriques en la naturalesa:

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1,6180339887498948...$$

El cos humà es creu que té una distribució a proporció d'aquest valor. Adolf Zeising, en el segle XIX, va realitzar la primera publicació analitzant aquest valor,

amb el títol *Nueva teoría de las proporciones del cuerpo humano, desarrolladas a partir de una ley morfológica básica hasta ahora desconocida, y que está presente en toda la naturaleza y el arte, acompañado por un resumen completo de los sistemas prevalentes*. En la Fig. 21 es mostra la distribució que va plantejar el Adolf Zeising.

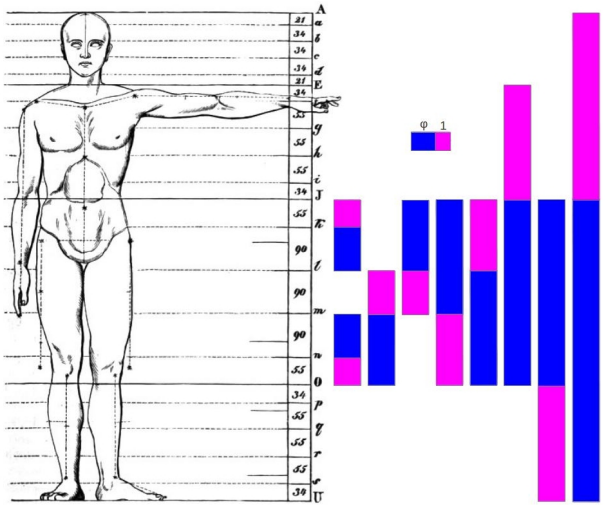

Figura 21: Golden Ratio ($\varphi$) [28]

Aquesta publicació s'ha anat ampliant al llarg del temps, quan el 1979 es va fer una nova publicació del Sr. T. Antony Davis i Rudolf Altevogt [28], on es realitzaven càlculs més concrets.

Amb el valor $\varphi$ es calcula el requadre de la Fig. 24 , el qual la seva altura representa l'altura del requadre verd pel valor Phi, i la seva amplada representa multiplicar per 2 l'amplada del requadre verd.

**Caracterització de la roba**

En el moment que es localitza una cara també s'extreuen els valors referents a la roba que porta. Aquests valors s'extreuen en forma de vector característic de 8 bits per canal, on es tracta d'una imatge de 3 canals, així que es crea un vector de 24 bits, el qual fa referencia a la roba que porta la persona.

Al tenir l'espai de colors concretat, es voldrà comprovar si aquest vector pot aportar més informació per poder classificar de manera més eficient els rostres, pel que es proposarà de concatenar el vector de la roba de 24 bits amb el vector característic dels rostres de 128 bits per a crear un vector característic de 132 bits, el qual es normalitzarà.

A part de crear un vector normalitzat de 132 bits, el qual s'utilitzarà amb els agrupaments, també es comprovarà la similitud dels vectors de roba, de 24 bits, amb la distància de bhatacharrya.

Utilitzem l'histograma d'espai de color **HSV** (Hue Saturation Value) com a descriptor de la roba. Aquest espai de color configurat amb tres canals que es descomponen en la tonalitat (canal H), saturació (canal S) i valor (canal V). La transformació de l'espai de color RGB a HSV no és lineal, tal com es pot mostrar en la següent equació:

$$H = \begin{cases} \textbf{no definido} & \textbf{si } MAX = MIN \\ 60^o \frac{G-B}{MAX-MIN} + 0^o & \textbf{si } MAX = R \textbf{ y } G \geq B \\ 60^o \frac{G-B}{MAX-MIN} + 360^o & \textbf{si } MAX = R \textbf{ y } G < B \\ 60^o \frac{G-R}{MAX-MIN} + 120^o & \textbf{si } MAX = G \\ 60^o \frac{R-G}{MAX-MIN} + 240^o & \textbf{si } MAX = B \end{cases}$$

$$S = \begin{cases} 0 & \textbf{si } MAX = 0 \\ 1 - \frac{MIN}{MAX} & \textbf{en otro caso} \end{cases}$$

$$V = MAX$$

Figura 22: Transformació de RGB a HSV

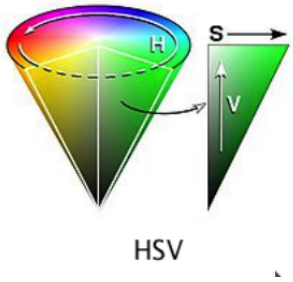


Figura 23: Exemples d'espais de color HSV

**La distància de Bhattacharyya**

Aquesta mesura s'utilitza per comparar els valors extrets de les zones de la roba, i s'ha volgut comprovar el seu funcionament pels resultats observats [24]. Al tractar-se d'un valor normalitzat, es pot concretar un valor de tall i amb aquest poder discriminar si la roba entre imatges és igual/similar o és diferent (Subsecció **??**).

La distància de Bhattacharyya mesura la similitud de dues distribucions de probabilitat discretes o continues. Està estretament relacionat amb el coeficient de Bhattacharyya, és una mesura de la quantitat de superposició entre dues mostres estadístiques o poblacions:

$$d_B(p,q) = \frac{1}{4} ln(\frac{1}{4}(\frac{\sigma_p^2}{\sigma_q^2} + \frac{\sigma_q^2}{\sigma_p^2} + 2)) = \sqrt{1 - \sum_i \sqrt{h_1(i) * h_2(i)}}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad \overline{x} = \frac{\sum x}{n}.$$

El coeficient es pot utilitzar per determinar la proximitat relativa de les dues mostres. S'utilitza per mesurar la separació de les classes en classificació i es considera més fiable que la distancia de Mahalanobis, ja que aquesta distància és un cas particular de la distància de Bhattacharyya quan les desviacions estàndards de les

dues classes són les mateixes [24]. En conseqüència, quan dues classes tenen mitjans similars, però diferents desviacions estàndard, la distància de Mahalanobis tendís a zero, mentre que la distància de Bhattacharyya creix dependent de la diferència entre les desviacions estàndards.

A part, la distància de Mahalanobis necessita una matriu de covariància, un factor que alenteix l'execució de l'aplicació, tal com es pot mostrar a la fórmula següent de la distància de Mahalanobis:

$$d_{Mahala}(x,y) = \sqrt{(x-y)\frac{1}{S}(x-y)} \qquad S \leftarrow \text{Matriu de covariància.}$$

**Exclusió de la pell**

Per a obtenir valors referits a sols els colors de la roba, s'ha generat un detector de pell per tal de localitzar els píxels on està ubicada la pell. Per tenir els valors del vector característic de la roba, s'han eliminat aquests píxels referents a la pell.

Per a poder determinar quins píxels estan vinculats amb la pell, s'han localitzat les zones del nas i les galtes. Amb els valors d'aquests píxels, hem definit el color de la pell de la persona. S'han comprovat diferents espais de colors per a obtenir els valors de colors més útils per la discriminació de la pell - els espais de Gray, RGB, YCrCb i HSV.

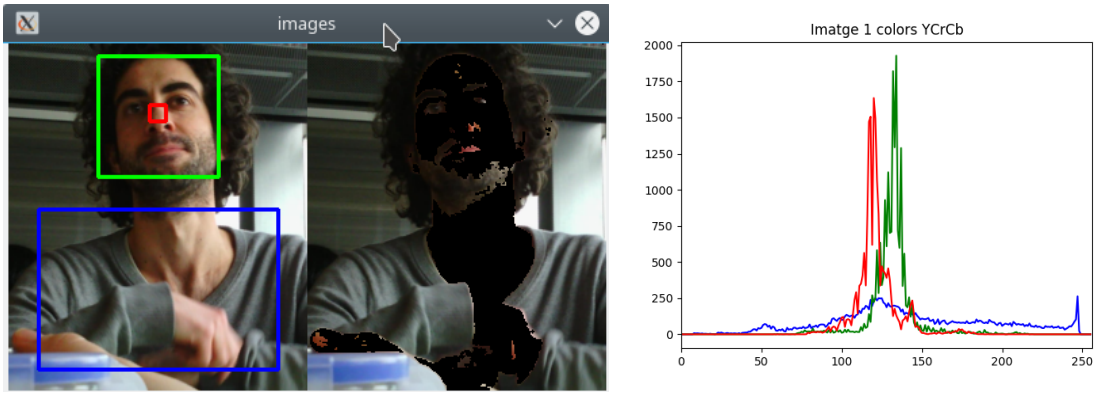


Figura 24: Detecció de Pell

En la Fig. 24 es pot observar com en el requadre verd es localitza la cara i gràcies a les dimensions de la cara es calculen les dimensions, gràcies al Golden Ratio, del requadre blau, el qual representa el cos on estan les peses de roba. També es pot observar el requadre vermell, el qual serveix per recopilar els valors de colors de la pell per posteriorment extreure-la, tal com es pot observar a la imatge de la dreta de l'esmentada figura, utilitzant l'espai de colors YCrCb.

La gràfica a la dreta, en la Fig. 24, representa els histogrames amb el que es farien les comparatives dels diferents espais de color, els quals es comprovaran en l'apartat d'experiments.

A pesar que aquest sistema no és totalment inflable, és una primera aproximació a incloure aquest tipus d'informació per a l'agrupament de cares. Diem que no és

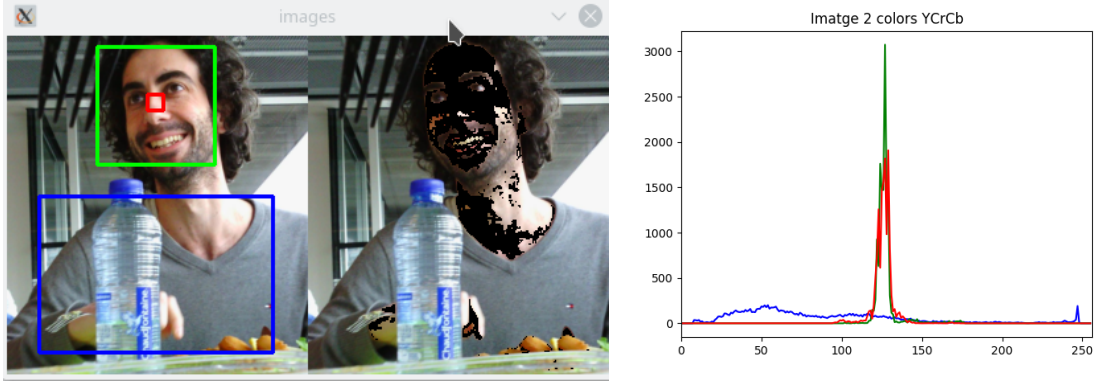


Figura 25: Detecció de Pell

fiable, ja que en determinades imatges no es poden eliminar tots els píxels vinculats amb la pell, tal com es mostra en la imatge de la Fig. 25.

## 3.3 Agrupament

Una vegada obtingudes les imatges retallades de les cares amb els seus vectors característics, s'apliquen mètodes d'agrupament - Çlustering". L'objectiu és obtenir diferents grups, on cada grup correspondria a una persona si la classificació fos correcta. En aquest projecte es plantegen tres mètodes d'agrupament diferents per tal de mostrar un estudi de la problemàtica. A continuació els detallarem, l'agrupament jeràrquic, basat en *MeanShift* i el *Spectral clustering*.

**Agrupament Jeràrquic**

Hi ha dos mètodes d'agrupaments jeràrquics: els d'aglomerat i els dissociatius. Els primers comencen l'anàlisi per unitats individuals i es van formant grups a partir d'aquestes unitats. Mentre que els segons, els mètodes dissociatius, són just el contrari, parteixen d'un grup inicial i a cada iteració es va disgregant en cerca de les unitats individuals.

En aquest projecte s'utilitzen els mètodes aglomerats, per tal de crear grups a partir de les unitats individuals, que en el nostre cas equival a cada exemple de cara. Per aquest motiu, s'utilitza la llibreria *Scipy* per *Python*.

El funcionament per crear aquests grups és molt simple: es crea una matriu de similitud, en aquesta matriu es calculen les distàncies entre les respectives unitats 'vectors característics' i el valor més baix és on es crearà la unió de les unitats.

Les dues unitats unides es contemplen com una sola unitat i es torna a crear la taula de similituds a la cerca de la pròxima unió amb el valor més baix.

El punt a tenir molt present és quin càlcul per crear la matriu de similitud s'utilitza, ja que l'agrupament jeràrquic disposa de 7 mètodes per calcular les distàncies de similitud.

En l'agrupament jeràrquic s'utilitza per defecte la distància Euclidiana, descrita

```
Z = linkage(x, method = metode ,metric='euclidean')
dn = dendrogram(Z,leaf_rotation=90.,leaf_font_size=10.,show_contracted=True,labels=labels)
```

Figura 26: Codi agrupament Jeràrquic

a continuació 3.3, per calcular les distancies entre els elements. Segons el mètode utilitzat els valors poden canviar, ja que cada mètode selecciona un tipus concret d'elements a calcular. En aquest treball proposam el mètode Average per a l'agrupament dels vectors de característiques.

Mètode **Average**: Es realitza una mitja ponderada de les distàncies de tots els elements d'un conjunt amb els elements de l'altre conjunt:

$$d(u,v) = \sum_{ij} \frac{d(u[i],v[j])}{(|u|*|v|)}$$

Per l'execució d'aquest agrupament es fa servir la funció: "Z = linkage(dades, method = 'mètode')", tal com es mostra en la Fig. 26, la qual crea la matriu de similitud segons el mètode de càlcul de les distàncies.

**La distància Euclidiana**

La distància Euclidiana o euclidià és la més utilitzada per moltes aplicacions, la qual es dedueix a partir del teorema de Pitàgores. Una funció no negativa utilitzada per calcular la distància entre dos punts, també serveix per definir la distancia entre dos punts en altres tipus d'espais de tres o més dimensions. S'utilitza per obtenir la longitud d'un segment definit per dos punts d'una recta, del pla o d'espai de major dimensió.

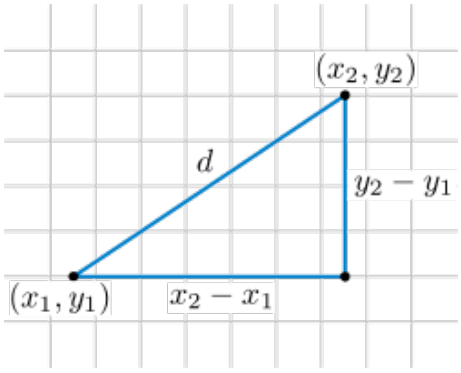


Figura 27: Distancia Euclidià

Per calcular la distància Euclidià es calcula amb la següent formula:

$$d_E(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_{ni} - y_{ni})^2}.$$

### Coeficient de Pearson

El coeficient de correlació de Pearson és una mesura de la relació lineal entre dos variables aleatòries quantitatives. A diferència de la covariància, la correlació de Pearson és independent de l'escala de mesura de les variables i es determina com:

$$r_{xy} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

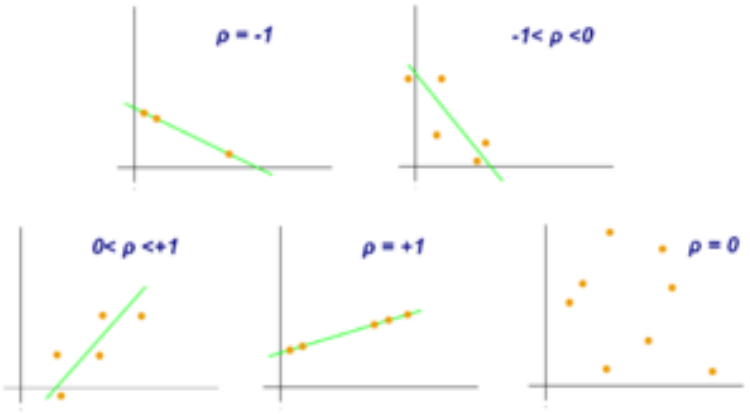


Figura 28: Coeficient de correlació de Pearson

De manera menys formal, es pot definir el coeficient de correlació de Pearson com un índex que pot utilitzar-se per mesurar el grau de relació entre dos variables sempre que siguin quantificables.

Aquest coeficient s'utilitza en aquest treball per poder agrupar d'una manera més precisa els vectors característics, ja que s'intenten agrupar per la similitud entre ells i aquest coeficient de correlació és molt útil. També, per a poder discriminar la roba, s'han comprovat els valors que representaven per l'histograma i es vol comprovar si es pot aplicar la correlació per poder discriminar si la roba és igual (similar) o diferent.

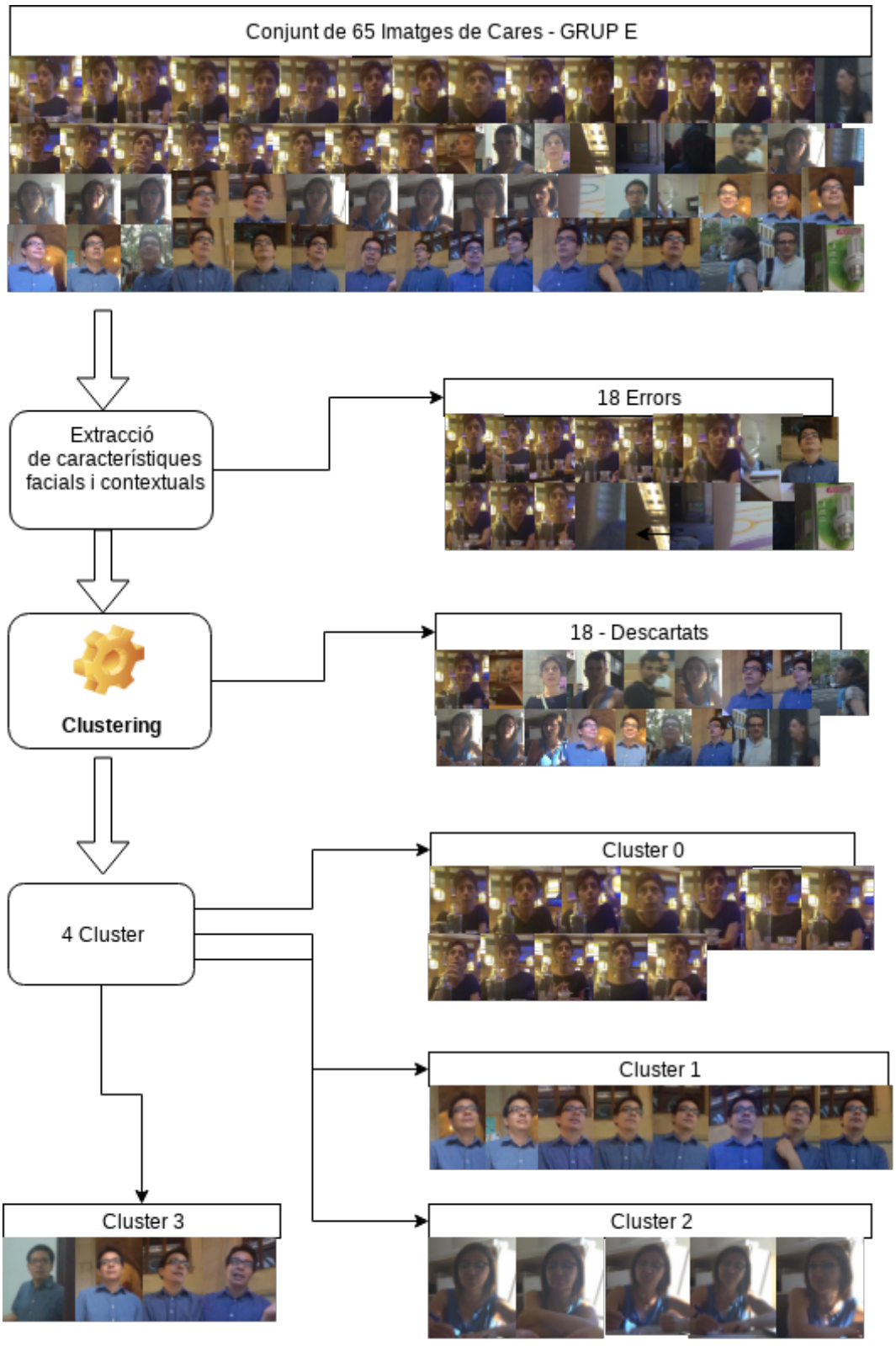


Figura 29: Funcionament del mètode plantejat

En la figura 29 es mostra la metodologia amb un resultat final una vegada s'ha aplicat un clustering jeràrquic amb el coef. de Pearson definint un valor de 0,70.

# 4 Validació

## 4.1 Dataset

En aquest apartat, s'analitzaran 6 sets de Dades concrets per tal de tenir la seva puntuació amb l'execució de l'algorisme Viola & Jones. A part, aquests sets de dades seran els que s'explotaran en els experiments posteriors, els sets de dades es reparteixen en els grups:

**Grup 0**: Conjunt inicial de 317 imatges amb una capacitat total de 115,6 Mb, on apareixen 20 persones diferents i consta de 377 cares.

**Grup A**: Conjunt inicial ampliat de 803 imatges amb una capacitat total de 264,2 Mb, on apareixen 37 persones diferents

**Grup B**: Conjunt de 751 imatges amb una capacitat total de 887,7 Mb, representen les imatges de l'Usuari 2 del dia 02-09-2016 i apareixen 8 persones diferents.

**Grup C**: Conjunt de 1302 imatges amb una capacitat total de 960,1 Mb, representen les imatges de l'Usuari 2 del dia 02-12-2016 i apareixen 13 persones diferents.

**Grup D**: Conjunt de 665 imatges amb una capacitat total de 339,1 Mb, representen les imatges de l'Usuari 5 del dia 02-06-2016 i apareixen 9 persones diferents.

**Grup E**: Conjunt de 760 imatges amb una capacitat total de 407,2 Mb, representen les imatges de l'Usuari 5 del dia 14-07-2016 i apareixen 3 persones diferents.

## 4.2 Mesures de validació

Per a realitzar les comprovacions i validacions dels experiments s'utilitza la mesura de control de **F-Score**, la qual es mostra en els resultats. Per poder obtenir el valor de F-Score, prèviament s'han d'obtenir les mètriques de Precision i Recall, les quals també es mostren en els experiments.

$$F_{Score} = \frac{Precision * Recall}{Precision + Recall}$$

La mètrica de **Precision** representa la proporció d'elements classificats correctament (True Positive) respecte als elements que són classificats de forma errònia (False Positive).

$$Precision = \frac{[TruesPositives]}{[TruesPositives] + [FalsesPositives]}$$

La mètrica de **Recall** o exhaustivitat representa els valors classificats correctament (True Postive) respecte als valors descartats erròniament (False Negative).

$$Recall = \frac{[TruesPositives]}{[TruePositives] + [FalsesNegatives]}$$

Per comparar l'eficàcia dels diferents mètodes proposats s'utilitzaven les mesures abans descrites, però s'ha de fer èmfasis que una premissa que es volia complir era que el sistema classifiques amb els menors errors possibles, és a dir, tenir la mètrica Precisió el més pròxim a 100%, per aquest motiu s'ha tingut molt present aquesta mètrica.

## 4.3 Comparació amb l'estat de l'art

La metodologia plantejada es comparada amb altres implementacions d'aquesta. Per això, implementam similars models, on es varien els paràmetres. Per exemple, utilitzam diferents sistemes d'agrupació i diferents distàncies per poder comprovar si hi ha un sistema amb millors resultats. També es comprovaren altres espais de colors envers els valors obtinguts de les característiques de la roba, on també es proposarà l'eficàcia d'analitzar la zona de la roba excloent o no la pell de la persona.

### 4.3.1 Mètodes d'agrupament

S'han implementat altres mètodes d'agrupament per tal de comparar el seu rendiment amb el del mètode que es proposa en aquest treball. Aquest mètodes son *MeanShift* i *SpectralClustering*.

**a) Agrupament Jeràrquic** Hem comparat el mètode proposat, jeràrquic amb distància average, amb els resultats obtinguts utilitzant les altres distancies.

Mètode **Single**: La distància que es calcula és des dels elements més pròxims entre els dos conjunts a comparar:

$$d(u, v) = min(dist(u[i], v[j]))$$

Mètode **Complete**: La distància que es calcula és des de l'element més allunyat entre els dos conjunts a comparar:

$$d(u, v) = max(dist(u[i], v[j]))$$

Mètode **Weighted**: Es realitza una mitja no ponderada de les distàncies entre els elements dels diferents conjunts:

$$d(u, v) = \frac{(dist(s,v)+dist(t,v))}{2}$$

Mètode **Centroid** : Es calcula la distància entre els centroides no ponderats dels conjunts d'elements.

$$dist(s, t) = ||C_s - C_t||_2$$

Mètode **Median**: Es calcula la distància entre els centroides ponderats dels conjunts d'elements.

Mètode **Ward**: Calcula la distància per cada dos conjunts per unir els dos que tenen els valors més baix i representen un menor increment en la suma del total:

$$d(u,v) = \sqrt{\frac{|v|+|s|}{T}d(v,s)^2 + \frac{|v|+|t|}{T}d(v,t)^2 + \frac{|v|}{T}d(s,t)^2}$$

**b) Agrupament MeanShift** Aquest agrupament implementa una idea molt simple que realitza qualsevol persona en analitzar les imatges. MeanShift considera un espai de característiques com una funció de densitat de probabilitat. Per implementar aquest tipus d'agrupament s'han utilitzat les llibreries "scikit-learn"per Python.

Si l'entrada d'informació és un conjunt de punts, MeanShift els considera com una mostra de la funció de probabilitat corresponent. Si hi ha regions denses, aquestes es corresponen amb un mode de la funció de densitat de probabilitat. Per cada punt de dades, l'algorisme fa desplaçar la finestra (bandwidth) fins al pic més proper de la funció de densitat de probabilitat del conjunt de dades. Per a cada punt de dades, el canvi de medi defineix una finestra al voltant d'ella i calcula la mitjana del punt de dades. Després es desplaça el centre de la finestra per a la mitjana i es repeteix l'algorisme fins que convergeix. Després de cada iteració, es pot considerar que la finestra es desplaça a una regió més densa del conjunt de dades. A alt nivell, podem especificar MeanShift de la següent manera:

- Preparar i normalitzar les dades per a poder ser processades, on la llibreria "scikit-learn"té la funció "preprocessing.scale('Dades')"per poder-ho realitzar, i

- Fixar una finestra (bandwidth) al voltant de cada punt de dades, on la llibreria "scikit-learn"té la funció "estimate_bandwidth" que fa una estimació del valor de la finestra.

Per l'execució d'aquest agrupament, és mostra en la Fig. 30, on hi ha el codi per fer el clustering.

**c) Agrupament Spectral**

Aquest tipus d'agrupament aplica una matriu de similitud amb la qual es realitza una reducció de dimensions gràcies a agrupar les dades. Es representa el clustering generant una partició del graf de forma de les línies que uneixen diferents grups que tinguin 'baix' pes (dissimiles entre si) i les línies que uneixen punts del mateix grup que el tinguin alt pes (semblants entre si).

Aquest sistema es configura amb les següents característiques basades en grafs:

- Un graf $G = (V, E)$ no direccional si $w_{ij} = w_{ji}$ , i conté pesos si $w_{ij} \geq 0$ per a tot $i, j$.

- Matriu d'Adjacència $W$ es la matriu amb els elements $w_{ij}$.

- Grau d'un vèrtex $v_i : d_i = \sum_{j=1}^{n} w_{ij}$.

```
#Es preparen les dades a escala per utilitzar MeanShift
X = preprocessing.scale(np.array(data))

# Compute clustering with MeanShift

# The following bandwidth can be automatically detected using
bandwidth = estimate_bandwidth(X,quantile=0.2, n_samples = 500)

ms = MeanShift(bandwidth = bandwidth, bin_seeding=False)
ms.fit(X)
labels = ms.labels_
cluster_centers = ms.cluster_centers_
```

Figura 30: Codi del Mean-Shift

- Matriu de greu $D$: es una matriu diagonal amb els elements $a_{ii} = d_i$.

- Adjacència entre A i B: $W(A, B) = \sum_{i \epsilon A, j \epsilon B} w_{ij}$.

- Dimensió de $A$: $|A|$ = número de vèrtex pertanyents a A; $vol(A) = \sum_{i \epsilon A} d_i$.

- $A$ connectat si per tot $i, j$ pertanyes a $A$, las línies o camí de línies que connecten $v_i$ amb $v_j$ estan incluses en $A$.

- A component connectada si és connectada i no existeixen connexions amb els vèrtexs de $\overline{A}$.

- Partició de $V : A_1 \cup A_2 \cup A_3 ... \cup A_k = VyA_i \cap A_j = \Phi$, per tot $i, j$.

Spectral clustering representa la similitud amb diferents grafs, per així modela les relacions de veïnat entre vèrtexs en diferents models. A continuació, es descriuen els diferents grafs de similitud.

**Grafs de veïnat-$\epsilon$**: Connecta dos vèrtexs $v_i, v_j$ si la distància és $d_{ij} \prec \epsilon$. És usual que aquest graf no tingui pesos associats.

**Grafs de k-veïns pròxims**: Connecta dos vèrtexs $v_i, v_j$ si $v_j$ està entre els k-veïns pròxims de $v_i$.

**Grafs de k-veïns mutus pròxims**: Connecta dos vèrtexs $v_i, v_j$ si $v_j$ està entre els k-veïns pròxims de $v_i$ i si $v_i$ està entre els k-veïns pròxims a $v_j$.

**Grafs totalment connectat**: Connectat dos vèrtexs $v_i, v_j$ si $s_{ij} \succ 0$ i els pesos $w_{ij}$ s'assignen segons $s_{ij}, s_{ij}$ hauria de modelar el veïnat local.

Per l'execució d'aquest agrupament, és mostra en la Fig. 31, on hi ha el codi per fer el clustering.

L'inconvenient d'aquest sistema d'agrupament és la necessitat d'indicar una $k$ per tal de realitzar tants agrupaments com $k$'s s'indiquin, per aquest motiu en la funció Spectral clustering hi ha la funció "elbow()"amb la que es calcula la $k$-optima aplicant el mètode del "colze".

```
100       gt = np.array(data)
101       #Metode elbow (del colze)
102       knum = km.elbow(data,aviso = False)
103       # Clustering Spectral
104       sc = SpectralClustering(knum,
105                               eigen_solver = None ,
106                               affinity = 'nearest_neighbors',
107                               n_init=500)
108       sc.fit(gt)
```

Figura 31: Codi Agrupament Spectral

### 4.3.2 Informació contextual - analysis de diferents espais de color

Comparam el rendiment quan s'utilitzen altres espais de color. Per això comparen diferents vectors característics de roba, sense concatenar altre tipus d'informaciò. Entre els vectors es realitzen les mesures comentades en els apartats 3.3 (Distància Euclidiana), 3.3 (Coeficient de Pearson) i 3.2.2 (Distancia de Bhattacharyya).

Hi ha una sèrie d'espais de color que tenen el seu propi sistema de coordenades de color, cada punt en el sistema de coordenades representa un color diferent. Existeix una àmplia varietat de models de colors, els quals posseeixen característiques que els fan útils en determinats tipus de problemes. En aquest treball fem una comparació utilitzant els diferents espais de colors com a descriptors dels píxels relacionats amb la roba.

Els espais de color que se solen aplicar en l'àmbit de visió per computació són [26]:

- **Gray** (Escala de grisos): És una escala empleada en la imatge digital, on el valor de cada píxel posseeix un valor equivalent a una graduació de gris. Les imatges representades amb aquest tipus d'espais estan compostes d'ombres de grisos.

Figura 32: Exemples d'espais de color Gray

- **RGB** (Red Green Blue): És el típic espai de color, el qual es troba en la majoria de dispositius. Està configurat per tres canals (vermell, verd i blau) on la lluminositat i la crominància no es troben per separat, cosa que dificulta l'anàlisi de la imatge perquè no pot separar el factor de la lluminositat del color. Normalment cada canal està format per 8 bits. No és un espai que es percebi de manera uniforme.

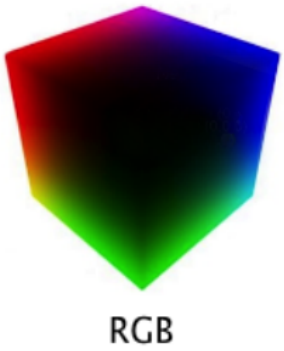


Figura 33: Exemples d'espais de color RGB

- **YCrCb** : Aquest espai de color es compon d'una component de lluminositat (Y) i dos components de color (Cb y Cr), que representa la crominància en blau i en vermell. La transformació d'espai RGB a YCrCb es representa amb la següent equació 4.1:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65,481 & 128,553 & 24,966 \\ -37,797 & -74,203 & 112 \\ 112 & -93,786 & -18,214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.1}$$

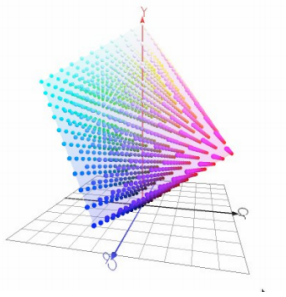

Figura 34: Exemples d'espais de color YCrCb

### 4.3.3 Vector Característic - Discriminador de Roba

A més, per a completar el set d'experiments, mostram els resultats obtinguts quan per a la classificació utilitzam exclusivament el vector característic de la roba. L'objectiu es el mateix que amb les configuracions anterior, el poder identificar si les zones de roba comparades son les mateixes. Per a portar a terme aquesta comparació s'utilitza la distància de Bhattacharrya.

# 5 Resultats

En aquesta secció exposem els resultats obtinguts pels diferents apartats. Hem dut a terme una extensa exploració de la performance dels diferents mètodes utilitzats per aquest project, i prèviament descrits en aquest treball. A més, reportem resultats obtinguts variant diferents paràmetres per a cada mètode descrit.

## 5.1 Detecció de cares per Viola & Jones

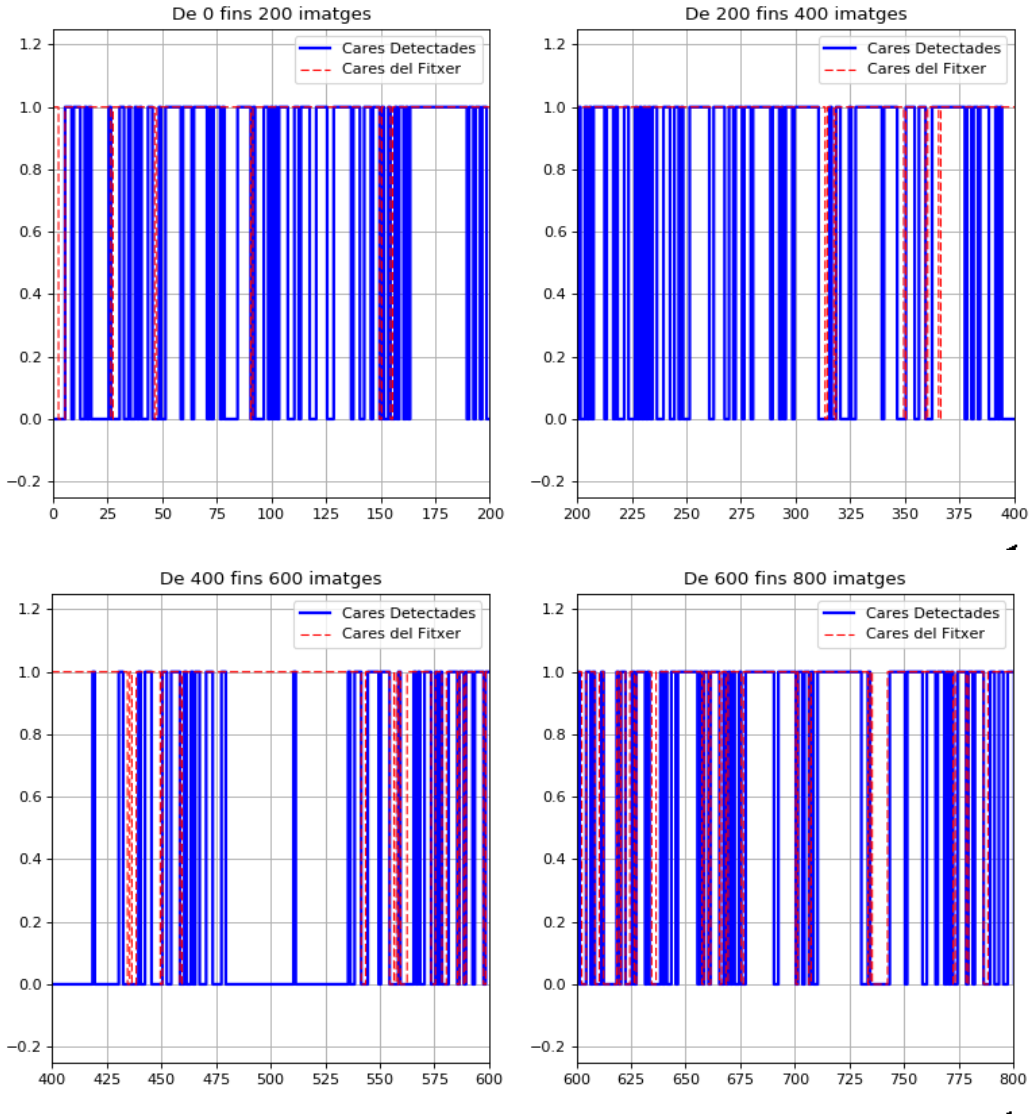


Figura 35: Mostrar localització de cares Grup A

Per la detecció de les cares amb el Algorisme Viola & Jones, pel qual s'utilitzen els filtres haardcascade_frontal_alt2.xml i haardcascade_profileface.xml, es comproven diversos paràmetres descrits a la metodologia per comprovar quina configuració té

una millor eficiència per les imatges egocèntriques, els resultats d'aquest punt es poden comprovar a les taules 4 i 5.

La Fig. 35 representa la localització de cares del Grup A d'imatges (desglossats a 5.1.5), on les línies blaves representen les cares detectades per Viola & Jones i les línies vermelles intermitents corresponen a la representació de les cares que apareixen en les imatges.

### 5.1.1 Velocitat d'execució de l'algorisme de Viola & Jones aplicat a les dades egocèntriques

Per comprovar la velocitat d'execució de l'algorisme de Viola & Jones, s'ha comprovat d'executar-lo amb diversos threads, on es mostren els resultats a la taula 4.

Les imatges amb les quals s'ha fet aquesta prova inicial, representen un conjunt de 233 imatges d'un set d'imatges públic de diverses fotografies familiars. Aquest conjunt d'imatges conté un total de 2028 cares. La capacitat total d'aquestes imatges és de 27,6 Mb amb unes dimensions de $1024x[500, 1100]$ píxels.

Taula 1: Execució Viola & Jones

| Paràmetres | | Deteccions | | Temps | | | Resultats | | |
|---|---|---|---|---|---|---|---|---|---|
| Scale Factor | Min Neig | Cares | False + | 1 Thread Seg. | 8 Threads Seg. | Guany | Precisio | Recall | F-Score |
| 1.1 | 5 | 1697 | 48 | 905,548 | 234,776 | 74,07% | 97,17% | 85,35% | 90,88% |
| | 8 | 1439 | 11 | 929,810 | 232,993 | 74,94% | 99,24% | 71,19% | 82,90% |
| | 11 | 1222 | 8 | 940,562 | 234,739 | 75,04% | 99,35% | 60,34% | 75,08% |
| | 14 | 1056 | 1 | 797,630 | 237,385 | 70,24% | 99,91% | 52,07% | 68,46% |
| 1.2 | 5 | 1340 | 9 | 518,262 | 139,651 | 73,05% | 99,33% | 66,22% | 79,46% |
| | 8 | 1050 | 2 | 342,998 | 138,437 | 59,64% | 99,81% | 51,78% | 68,18% |
| | 11 | 847 | 0 | 383,992 | 138,920 | 63,82% | 100,00% | 41,77% | 58,92% |
| | 14 | 699 | 0 | 340,571 | 140,108 | 58,86% | 100,00% | 34,47% | 51,27% |
| 1.3 | 5 | 1044 | 4 | 247,040 | 102,517 | 58,50% | 99,62% | 51,49% | 67,89% |
| | 8 | 782 | 0 | 246,123 | 101,091 | 58,93% | 100,00% | 38,56% | 55,66% |
| | 11 | 594 | 0 | 247,440 | 100,413 | 59,42% | 100,00% | 29,29% | 45,31% |
| | 14 | 412 | 0 | 246,803 | 100,767 | 59,17% | 100,00% | 20,32% | 33,77% |
| 1.4 | 5 | 584 | 1 | 180,782 | 75,065 | 58,48% | 99,83% | 28,78% | 44,67% |
| | 8 | 284 | 0 | 180,917 | 73,883 | 59,16% | 100,00% | 14,00% | 24,57% |
| | 11 | 108 | 0 | 184,416 | 74,185 | 59,77% | 100,00% | 5,33% | 10,11% |
| | 14 | 41 | 0 | 194,433 | 74,073 | 61,90% | 100,00% | 2,02% | 3,96% |

Per a millorar el rendiment d'aquest algorisme, s'ha paral·lelitzat la seva excussió, ja que s'ha comprovat que la millora d'executar l'algoritme amb 1 Core a 8 Cores és entre el 50% al 70% de guany, tal com es pot demostrar en la taula 4 de l'apartat de resultats.

### 5.1.2 Paràmetres de l'algorisme de Viola & Jones

En aquests experiment es comprovarà la millor configuració de paràmetres per fixar en les posteriors proves, on el resultat d'aquest experiment es mostra a la taula 5.

Per realitzar aquest experiment, s'ha utilitzat un conjunt inicial, Grup 0, de 317 imatges egocèntriques amb una capacitat total de 115,6 Mb, on apareixen 20 persones diferents i consta de 377 cares.

Taula 2: Resultats detecció de cares en imatges egocèntriques

| Paràmetres | | Resultats Viola | | Temps | | Rendiment Viola | | |
|---|---|---|---|---|---|---|---|---|
| Scale Factor | Min neighbors | Faces Detect | False Positive | Segons | Minuts | Precisio | Recall | F-Score |
| 1.1 | 5 | 373 | 117 | 1070,789 | 17,846 | 68,63% | 67,90% | 68,27% |
| | 8 | 245 | 28 | 973,091 | 16,218 | 88,57% | 57,56% | 69,77% |
| | 11 | 220 | 18 | 979,734 | 16,329 | 91,82% | 53,58% | 67,67% |
| | 14 | 200 | 4 | 1003,013 | 16,717 | 98,00% | 51,99% | 67,94% |
| | 17 | 170 | 2 | 986,017 | 16,434 | 98,82% | 44,56% | 61,43% |
| | 20 | 152 | 1 | 982,558 | 16,376 | 99,34% | 40,05% | 57,09% |
| | 23 | 140 | 1 | 979,944 | 16,332 | 99,29% | 36,87% | 53,77% |
| | 26 | 121 | 0 | 1040,930 | 17,349 | 100,00% | 32,10% | 48,59% |
| | 29 | 110 | 0 | 1031,473 | 17,191 | 100,00% | 29,18% | 45,17% |
| 1.2 | 5 | 240 | 24 | 579,833 | 9,664 | 90,00% | 57,29% | 70,02% |
| | 8 | 174 | 4 | 565,575 | 9,426 | 97,70% | 45,09% | 61,71% |
| | 11 | 141 | 0 | 572,530 | 9,542 | 100,00% | 37,40% | 54,44% |
| | 14 | 120 | 0 | 571,652 | 9,528 | 100,00% | 31,83% | 48,29% |
| | 17 | 94 | 0 | 565,504 | 9,425 | 100,00% | 24,93% | 39,92% |
| | 20 | 74 | 0 | 569,707 | 9,495 | 100,00% | 19,63% | 32,82% |
| 1.3 | 5 | 194 | 9 | 441,182 | 7,353 | 95,36% | 49,07% | 64,80% |
| | 8 | 142 | 0 | 437,514 | 7,292 | 100,00% | 37,67% | 54,72% |
| | 11 | 100 | 0 | 438,197 | 7,303 | 100,00% | 26,53% | 41,93% |
| | 14 | 70 | 0 | 411,005 | 6,850 | 100,00% | 18,57% | 31,32% |
| | 17 | 35 | 0 | 413,783 | 6,896 | 100,00% | 9,28% | 16,99% |
| | 20 | 16 | 0 | 432,420 | 7,207 | 100,00% | 4,24% | 8,14% |
| 1.4 | 5 | 125 | 1 | 300,334 | 5,006 | 99,20% | 32,89% | 49,40% |
| | 8 | 63 | 0 | 302,570 | 5,043 | 100,00% | 16,71% | 28,64% |
| | 11 | 20 | 0 | 319,269 | 5,321 | 100,00% | 5,31% | 10,08% |
| | 14 | 2 | 0 | 313,762 | 5,229 | 100,00% | 0,53% | 1,06% |
| | 17 | 0 | 0 | | 0 | 0,00% | 0,00% | 0,00% |
| | 20 | 0 | 0 | | 0 | 0,00% | 0,00% | 0,00% |

Com es pot observar en la taula 5, la millor configuració de paràmetres és **scale_Factor 1.2 i min_neighbors 5**, el qual ha obtingut un resultat de F-Score de **70,02%**.

### 5.1.3 Anàlisis de totes les dades amb Viola & Jones

En aquest experiment, es comprovaran els millors paràmetres segons el resultat de F-Score, es realitza una prova als Grups del A a l'E, abans descrits, per tenir els

resultats amb els paràmetres Scale Factor 1.2 min_neighbors 5, però en aquest punt es limiten els valors de minSize i maxSize, els quals són 5% i 35% respectivament de l'amplada de la imatge.

Taula 3: Resultats Viola & Jones Grups [A,B]

| Grup | Faces Detect | Temps Seg. | Precision | Recall | F-Score |
|---|---|---|---|---|---|
| A | 533 | 170,45 | 99,54% | 58,50% | 73,69% |
| B | 124 | 183,58 | 71,05% | 60,45% | 65,32% |
| C | 180 | 177,46 | 83,44% | 45,96% | 59,28% |
| D | 135 | 152,68 | 80,34% | 87,04% | 83,56% |
| E | 65 | 164,44 | 80,00% | 31,33% | 45,02% |

Al limitar els valor de minSize i maxSize, es pot observar un augment de la velocitat d'execució molt considerable amb uns resultats molt òptims, tal com es pot observar a la taula 6.

Per la detecció de les cares amb el Algorisme Viola & Jones, pel qual s'utilitzen els filtres haardcascade_frontal_alt2.xml i haardcascade_profileface.xml, es comproven diversos paràmetres descrits a la metodologia per comprovar quina configuració té una millor eficiència per les imatges egocèntriques, els resultats d'aquest punt es poden comprovar a les taules 4 i 5.

La Fig. 35 representa la localització de cares del Grup A d'imatges (desglossats a 5.1.5), on les línies blaves representen les cares detectades per Viola & Jones i les línies vermelles intermitents corresponen a la representació de les cares que apareixen en les imatges.

### 5.1.4 Privatització d'imatges egocèntriques

En aquest apartat es detallaran els experiments realitzats en Privatització d'imatges egocèntriques.

#### 5.1.4.1 Velocitat d'execució de l'algorisme de Viola & Jones aplicat a les dades egocèntriques

Per comprovar la velocitat d'execució de l'algorisme de Viola & Jones, s'ha comprovat d'executar-lo amb diversos threads, on es mostren els resultats a la taula 4.

Les imatges amb les quals s'ha fet aquesta prova inicial, representen un conjunt de 233 imatges d'un set d'imatges públic de diverses fotografies familiars. Aquest conjunt d'imatges conté un total de 2028 cares. La capacitat total d'aquestes imatges és de 27,6 Mb amb unes dimensions de $1024x[500, 1100]$ píxels.

Taula 4: Execució Viola & Jones

| Paràmetres | | Deteccions | | Temps | | | Resultats | | |
|---|---|---|---|---|---|---|---|---|---|
| Scale Factor | Min Neig | Cares | False + | 1 Thread Seg. | 8 Threads Seg. | Guany | Precisio | Recall | F-Score |
| 1.1 | 5 | 1697 | 48 | 905,548 | 234,776 | 74,07% | 97,17% | 85,35% | 90,88% |
| | 8 | 1439 | 11 | 929,810 | 232,993 | 74,94% | 99,24% | 71,19% | 82,90% |
| | 11 | 1222 | 8 | 940,562 | 234,739 | 75,04% | 99,35% | 60,34% | 75,08% |
| | 14 | 1056 | 1 | 797,630 | 237,385 | 70,24% | 99,91% | 52,07% | 68,46% |
| 1.2 | 5 | 1340 | 9 | 518,262 | 139,651 | 73,05% | 99,33% | 66,22% | 79,46% |
| | 8 | 1050 | 2 | 342,998 | 138,437 | 59,64% | 99,81% | 51,78% | 68,18% |
| | 11 | 847 | 0 | 383,992 | 138,920 | 63,82% | 100,00% | 41,77% | 58,92% |
| | 14 | 699 | 0 | 340,571 | 140,108 | 58,86% | 100,00% | 34,47% | 51,27% |
| 1.3 | 5 | 1044 | 4 | 247,040 | 102,517 | 58,50% | 99,62% | 51,49% | 67,89% |
| | 8 | 782 | 0 | 246,123 | 101,091 | 58,93% | 100,00% | 38,56% | 55,66% |
| | 11 | 594 | 0 | 247,440 | 100,413 | 59,42% | 100,00% | 29,29% | 45,31% |
| | 14 | 412 | 0 | 246,803 | 100,767 | 59,17% | 100,00% | 20,32% | 33,77% |
| 1.4 | 5 | 584 | 1 | 180,782 | 75,065 | 58,48% | 99,83% | 28,78% | 44,67% |
| | 8 | 284 | 0 | 180,917 | 73,883 | 59,16% | 100,00% | 14,00% | 24,57% |
| | 11 | 108 | 0 | 184,416 | 74,185 | 59,77% | 100,00% | 5,33% | 10,11% |
| | 14 | 41 | 0 | 194,433 | 74,073 | 61,90% | 100,00% | 2,02% | 3,96% |

Per a millorar el rendiment d'aquest algorisme, s'ha paral·lelitzat la seva excussió, ja que s'ha comprovat que la millora d'executar l'algoritme amb 1 Core a 8 Cores és entre el 50% al 70% de guany, tal com es pot demostrar en la taula 4 de l'apartat de resultats.

#### 5.1.4.2 Paràmetres de l'algorisme de Viola & Jones

En aquests experiment es comprovarà la millor configuració de paràmetres per fixar en les posteriors proves, on el resultat d'aquest experiment es mostra a la taula 5.

Per realitzar aquest experiment, s'ha utilitzat un conjunt inicial, Grup 0, de 317 imatges egocèntriques amb una capacitat total de 115,6 Mb, on apareixen 20 persones diferents i consta de 377 cares.

Taula 5: Resultats detecció de cares en imatges egocèntriques

| Paràmetres | | Resultats Viola | | Temps | | Rendiment Viola | | |
|---|---|---|---|---|---|---|---|---|
| Scale Factor | Min neighbors | Faces Detect | False Positive | Segons | Minuts | Precisio | Recall | F-Score |
| 1.1 | 5 | 373 | 117 | 1070,789 | 17,846 | 68,63% | 67,90% | 68,27% |
| | 8 | 245 | 28 | 973,091 | 16,218 | 88,57% | 57,56% | 69,77% |
| | 11 | 220 | 18 | 979,734 | 16,329 | 91,82% | 53,58% | 67,67% |
| | 14 | 200 | 4 | 1003,013 | 16,717 | 98,00% | 51,99% | 67,94% |
| | 17 | 170 | 2 | 986,017 | 16,434 | 98,82% | 44,56% | 61,43% |
| | 20 | 152 | 1 | 982,558 | 16,376 | 99,34% | 40,05% | 57,09% |
| | 23 | 140 | 1 | 979,944 | 16,332 | 99,29% | 36,87% | 53,77% |
| | 26 | 121 | 0 | 1040,930 | 17,349 | 100,00% | 32,10% | 48,59% |
| | 29 | 110 | 0 | 1031,473 | 17,191 | 100,00% | 29,18% | 45,17% |
| 1.2 | 5 | 240 | 24 | 579,833 | 9,664 | 90,00% | 57,29% | 70,02% |
| | 8 | 174 | 4 | 565,575 | 9,426 | 97,70% | 45,09% | 61,71% |
| | 11 | 141 | 0 | 572,530 | 9,542 | 100,00% | 37,40% | 54,44% |
| | 14 | 120 | 0 | 571,652 | 9,528 | 100,00% | 31,83% | 48,29% |
| | 17 | 94 | 0 | 565,504 | 9,425 | 100,00% | 24,93% | 39,92% |
| | 20 | 74 | 0 | 569,707 | 9,495 | 100,00% | 19,63% | 32,82% |
| 1.3 | 5 | 194 | 9 | 441,182 | 7,353 | 95,36% | 49,07% | 64,80% |
| | 8 | 142 | 0 | 437,514 | 7,292 | 100,00% | 37,67% | 54,72% |
| | 11 | 100 | 0 | 438,197 | 7,303 | 100,00% | 26,53% | 41,93% |
| | 14 | 70 | 0 | 411,005 | 6,850 | 100,00% | 18,57% | 31,32% |
| | 17 | 35 | 0 | 413,783 | 6,896 | 100,00% | 9,28% | 16,99% |
| | 20 | 16 | 0 | 432,420 | 7,207 | 100,00% | 4,24% | 8,14% |
| 1.4 | 5 | 125 | 1 | 300,334 | 5,006 | 99,20% | 32,89% | 49,40% |
| | 8 | 63 | 0 | 302,570 | 5,043 | 100,00% | 16,71% | 28,64% |
| | 11 | 20 | 0 | 319,269 | 5,321 | 100,00% | 5,31% | 10,08% |
| | 14 | 2 | 0 | 313,762 | 5,229 | 100,00% | 0,53% | 1,06% |
| | 17 | 0 | 0 | | 0 | 0,00% | 0,00% | 0,00% |
| | 20 | 0 | 0 | | 0 | 0,00% | 0,00% | 0,00% |

Com es pot observar en la taula 5, la millor configuració de paràmetres és **scale_Factor 1.2 i min_neighbors 5**, el qual ha obtingut un resultat de F-Score de **70,02%**.

### 5.1.5 Anàlisis de totes les dades amb Viola & Jones

En aquest experiment, es comprovaran els millors paràmetres segons el resultat de F-Score, es realitza una prova als Grups del A a l'E, abans descrits, per tenir els resultats amb els paràmetres Scale Factor 1.2 min_neighbors 5, però en aquest punt es limiten els valors de minSize i maxSize, els quals són 5% i 35% respectivament de l'amplada de la imatge.

**Resultats:**

Taula 6: Resultats Viola & Jones Grups [A,B]

| Grup | Faces Detect | Temps Seg. | Precision | Recall | F-Score |
|---|---|---|---|---|---|
| A | 533 | 170,45 | 99,54% | 58,50% | 73,69% |
| B | 124 | 183,58 | 71,05% | 60,45% | 65,32% |
| C | 180 | 177,46 | 83,44% | 45,96% | 59,28% |
| D | 135 | 152,68 | 80,34% | 87,04% | 83,56% |
| E | 65 | 164,44 | 80,00% | 31,33% | 45,02% |

Al limitar els valor de minSize i maxSize, es pot observar un augment de la velocitat d'execució molt considerable amb uns resultats molt òptims, tal com es pot observar a la taula 6.

## 5.2 Agrupament

Amb els retalls de les imatges, on es troben els possibles rostres, generats en els experiments anteriors de Viola & Jones s'han realitzat les proves en aquest apartat, ja que les imatges de cada grup seran agrupades segons els seus vectors característics obtinguts de la llibreria OpenFace [20].

Els paràmetres utilitzats per les proves fetes amb els agrupaments seran: Factor 1.2, min neighbors 5, min size 5% i maxSize 35%, amb els filtres haardcascade_frontal_alt2.xml i haardcascade_profileface.xml

### 5.2.1 Agrupament Jerarquic - sense tall

Per analitzar quin mètode és el més efectiu, es comproven tots els mètodes en cada grup. L'objectiu és poder determinar si hi ha algun mètode que agrupi millor aquest tipus de dades, per a poder tindre una bona precisió a l'hora d'agrupar rostres i que aquest agrupin per similitud.

En aquest experiment les entrades d'informació són el conjunt de vectors característics dels rostres, extrets per OpenFace. També concretar el concepte d'error, segons es mostren en les taules, ja que aquest concepte indica que una imatge retallada amb un suposat rostre no s'ha pogut extreure el seu vector característic. En el cas que els clústers estiguessin formats per una sola imatge, aquest es consideren

com imatges descartades, pel que en principi no tindrien cap similitud amb altres imatges.

Les agrupacions es crearan sense utilitzar un valor de tall, així que cada clúster serà fruit del que hagi concretat el sistema jeràrquic. En aquestes agrupacions cal tenir present que no es parteixen amb totes les possibles cares si l'algorisme anterior (Viola & Jones) ha tingut un Recall inferior a 100%, això vol dir que el valor total de possibles cares és el valor inicial del conjunt d'imatges i si les imatges retallades són inferiors a les reals, el Recall mai serà del 100%. Com a màxim el Recall esperat serà el que es mostra a la taula 6.

Taula 7: Agrupament Jeràrquic sobre Grup A sense tall

| Mètode | Errors | Descarts | Clústers | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|
| Single | 54 | 81 | 19 | 45,21% | 33,11% | 38,22 |
| Complete | 54 | 0 | 24 | 15,26% | 26,19% | 19,28% |
| Ward | 54 | 0 | 3 | 22,59% | 23,20% | 22,89% |
| Average | 54 | 4 | 29 | 43,95 | 30,10% | 35,73% |
| Weighted | 54 | 0 | 34 | 18,42% | 31,39% | 23,21% |
| Centroid | 54 | 44 | 14 | 26,74% | 28,19% | 27,44% |
| Median | 54 | 30 | 10 | 34,88% | 28,37% | 31,29% |

Com es pot observar en la taula 7, els mètodes a destacar són el **Single** i el **Average** amb una precisió de més del 40%, els valors més alt. També s'ha de tenir present que el mètode del enllaçament singular ha descartat més rostre que cap altre mètode, pel que si queden menys rostres per organitzar la precisió pot ser més alta. Igualment es comprovaran tots els mètodes en els grups restants.

Taula 8: Agrupament Jeràrquic Grup B sense tall

| Mètode | Errors | Descarts | Clústers | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|
| Single | 40 | 6 | 2 | 85,25% | 39,10% | 56,61% |
| Complete | 40 | 1 | 5 | 94,37% | 45,58% | 61,47% |
| Ward | 40 | 0 | 2 | 79,71% | 41,04% | 54,19% |
| Average | 40 | 1 | 3 | 82,35% | 41,18% | 54,90% |
| Weighted | 40 | 6 | 3 | 96,83% | 43,88% | 60,40% |
| Centroid | 40 | 2 | 2 | 80,88% | 41,04% | 54,46% |
| Median | 40 | 8 | 3 | 98,39% | 43,57% | 60,40% |

Com es pot comprovar a la taula 8, el mètode a destacar és el **Weighted** amb el valor més alt, de 96,83% de precisió. També es destaca que la resta de valors han estat molt millors pels resultats del Grup A. On una possible raó, ja que tots els mètodes donen una precisió baixa, podria ser la quantitat d'imatges a agrupar, ja que grup A conte més imatges de rostres que els altres grups. També podria succeir que la qualitat de la definició del rostre fes aquesta mala agrupació, ja que el grup A és el que té més imatges de persones que es trobaven de fons.

Com es pot comprovar a la taula 9, el mètode a destacar és el **Average** amb

Taula 9: Agrupament Jeràrquic Grup C sense tall

| Mètode | Errors | Descarts | Clústers | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|
| Single | 75 | 13 | 11 | 52,81% | 18,29% | 27,17% |
| Complete | 75 | 0 | 12 | 53,33% | 21,62% | 30,77% |
| Ward | 75 | 0 | 3 | 14,00% | 6,48% | 8,86% |
| Average | 75 | 1 | 14 | 83,16% | 34,39% | 48,66% |
| Weighted | 75 | 2 | 11 | 65,38% | 18,40% | 28,71% |
| Centroid | 75 | 15 | 10 | 66,28% | 21,19% | 32,11% |
| Median | 75 | 16 | 13 | 66,36% | 25,00% | 36,32% |

el valor més alt, de 83,16% de precisió. També, destacar que és el mètode que ha agrupat de manera més precisa i sols descartant una imatge.

Taula 10: Agrupament Jeràrquic D sense tall

| Mètode | Errors | Descarts | Clústers | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|
| Single | 15 | 17 | 2 | 31,91% | 50,00% | 38,96% |
| Complete | 15 | 2 | 7 | 85,85% | 79,13% | 82,35% |
| Ward | 15 | 0 | 4 | 87,50% | 63,77% | 73,77% |
| Average | 15 | 9 | 7 | 95,92% | 77,69% | 85,84% |
| Weighted | 15 | 9 | 8 | 95,74% | 76,27% | 84,91% |
| Centroid | 15 | 18 | 4 | 100,00% | 75,41% | 85,98% |
| Median | 15 | 18 | 4 | 86,96% | 71,43% | 78,43% |

Com es pot comprovar a la taula 10, el mètode a destacar és el **Centroid** amb un valor de precisió del 100,00%. També destacar que a excepció del mètode Single tots han tingut molt bons resultats.

Taula 11: Agrupament Jeràrquic E sense tall

| Metode | Errors | Descarts | Clústers | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|
| Single | 18 | 9 | 4 | 95,24% | 21,69% | 35,34% |
| Complete | 18 | 3 | 7 | 90,20% | 23,81% | 37,67% |
| Ward | 18 | 0 | 3 | 85,11% | 23,39% | 36,69% |
| Average | 18 | 8 | 6 | 95,83% | 22,29% | 36,17% |
| Weighted | 18 | 0 | 8 | 85,45% | 24,12% | 37,62% |
| Centroid | 18 | 8 | 3 | 95,24% | 22,02% | 35,77% |
| Median | 18 | 6 | 2 | 56,41% | 14,57% | 23,16% |

Com es pot comprovar a la taula 11, el mètode a destacar és el **Average** amb el valor més alt de precisió, de 95,83%. Però s'ha d'indicar els bons resultats de la precisió, excepte el mètode Median.

Com s'ha comprovat amb els resultats dels diferents grups, el mètode més òptim no està concret, ja que segons les dades que es processen potser òptim qualsevol, per aquest motiu s'ha generat la següent taula 12 en la qual es fa una mitjana de tots els grups per cada mètode, així es pot concretar un mètode a utilitzar.

Taula 12: Mitjana dels mètodes d'agrupament Jeràrquic

| Metode | Precision | Recall | F-Score |
|---|---|---|---|
| Single | 68,47% | 32,43% | 44,01% |
| Complete | 67,80% | 39,27% | 49,80% |
| Ward | 57,78% | 31,58% | 40,84% |
| Average | 80,25% | 41,13% | 54,39% |
| Weighted | 72,36% | 38,81% | 50,52% |
| Centroid | 73,83% | 37,57% | 49,80% |
| Median | 68,60% | 36,59% | 47,72% |

En la taula 12 es mostren les mitjanes de tots els mètodes segons els resultats dels diferents grups de proves per tal de poder comprovar quin dels mètodes és el més òptim per l'agrupament de cares. Com es pot observar el mètode amb millors resultats és el **Average** amb un resultat de precisió de 80,25% i tenint el valor F-Score més alt.

### 5.2.2 Agrupament Jerarquic - amb valor de tall

Amb els resultats de la taula 12, se selecciona el mètode **Average** per experimentar amb l'agrupament jeràrquic, quan aquest se li aplica un valor de tall.

El valor de tall representa la distància de similitud entre les comparacions que fan els vectors de la cara. Ja que el mètode *average* no deixa de ser una mitjana ponderada i la comparació entre rostre es fa per la distància Euclidiana de cada vector a comparar, s'han escollit els valors de tall com a: 0.80, 0.90 i 1.00.

L'experiment es realitza amb tots els grups de proves abans explicats al principi de l'apartat i es podrà observa la quantitat de clúster que genera cada tall.

**Resultats:**

Amb aquest sistema es dóna molta dispersió de les imatges, com es pot veure a la taula 13, on moltes imatges iguals surten en diferents clústers. També, s'ha de posar de manifest que els valors de tall pel mètode average varien poc, ja que són mitjanes, perquè en el cas d'utilitzar el mètode single no es podria determinar un valor de tall concret, ja que els valors canvien segons els mínims dels elements del vector.

#### 5.2.2.1 Agrupament amb Pearson

Es vol millorar la precisió de l'agrupament, en poques paraules, que les imatges que agrupin representin a la mateixa persona per cada clúster. Per aquest objectiu s'ha plantejat el següent experiment.

Una vegada es generin els clústers per mitjà de l'agrupament,, es pretén fer una anàlisi d'aquest per comprovar si hi ha vectors que no corresponguin a la persona que té la majoria de vectors en el clúster.

Taula 13: Agrupament jeràrquic mètode *average* amb talls

| Grup | Errors | Tall | Descartes | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 54 | 0,90 | 16 | 60 | 21,23% | 29,78% | 24,79% |
| | | 1,00 | 2 | 29 | 24,10% | 35,35% | 28,66% |
| | | 1,10 | 1 | 16 | 31,75% | 39,35% | 35,11% |
| B | 40 | 0,90 | 9 | 3 | 85,88% | 42,86% | 57,18% |
| | | 1,00 | 8 | 2 | 75,64% | 40,60% | 52,84% |
| | | 1,10 | 3 | 3 | 72,62% | 41,79% | 53,05% |
| C | 75 | 0,90 | 4 | 14 | 51,02% | 19,38% | 28,09% |
| | | 1,00 | 2 | 11 | 57,58% | 21,84% | 31,67% |
| | | 1,10 | 0 | 7 | 48,48% | 19,20% | 27,51% |
| D | 15 | 0,90 | 12 | 8 | 96,77% | 75,63% | 84,91% |
| | | 1,00 | 6 | 7 | 98,35% | 61,76% | 75,87% |
| | | 1,10 | 2 | 5 | 42,72% | 64,71% | 51,46% |
| E | 18 | 0,90 | 8 | 6 | 100,00% | 22,29% | 36,45% |
| | | 1,00 | 3 | 6 | 88,00% | 23,53% | 37,13% |
| | | 1,10 | 2 | 5 | 85,11% | 23,39% | 36,69% |

S'aplicarà una agrupació jeràrquica sense talls pel mètode d'average i una vegada creats els clústers, es comprovarà la consistència d'aquest per mitjà del coeficient de correlació de Pearson (3.3). Aquest coeficient comprovarà si hi ha relació entre els elements del clúster i en cas de localitzar elements no correlatius, els treu del clúster perquè es torni a generar un altre agrupament amb tots els elements extret o en cas de no donar resultat etiquetar-los com a elements a descartar.

La correlació es fa per mitjà dels vectors característics obtinguts amb la funció d'OpenFace que genera un vector de 128 valors. Es comprovarà que els elements que contingui el clúster tinguin un valor pròxim a 1, que indica màxima similitud, però també s'aplicarà un valor mínim, el qual si un element no arriba, aquest serà tret del clúster. Es proposa comprovar l'eficàcia de 3 valors, els quals són 0.70, 0.75 i 0.80.

Com es mostra a la taula 14, es pot observar com la precisió augmenta de manera significativa segons el valor de comprovació del coeficient de Pearson. En contra partida, en augmentar el valor mínim del coeficient de Pearson, s'augmenta el nombre d'elements que acaben descartats i això significa que les imatges que no són similars queden descartades provocant que el valor de Recall sigui inferior, ja que incrementa el valor de falsos negatius.

També, observar que amb més precisió el nombre de clústers es redueix a causa de l'expulsió dels elements menys similars, és a dir, inferior al valor mínim indicat pel coeficient de Pearson. Això provoca que hi hagi menys similitud, però que aquestes siguin més fiables.

Taula 14: Agrupament Jeràrquic sense tall amb Pearson

| Grup | Erros | Pearson | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 54 | 0,70 | 193 | 32 | 35,86% | 22,30% | 27,43% |
| | | 0,75 | 314 | 13 | 67,78% | 17,30% | 27,57% |
| | | 0,80 | 377 | 6 | 82,35% | 11,62% | 20,36% |
| B | 40 | 0,70 | 14 | 3 | 87,67% | 41,55% | 56,38% |
| | | 0,75 | 17 | 4 | 100,00% | 42,14% | 59,30% |
| | | 0,80 | 42 | 2 | 100,00% | 31,34% | 47,73% |
| C | 75 | 0,70 | 33 | 14 | 87,00% | 14,12% | 24,30% |
| | | 0,75 | 64 | 11 | 88,68% | 7,97% | 14,63% |
| | | 0,80 | 90 | 5 | 95,24% | 3,58% | 6,90% |
| D | 15 | 0,70 | 24 | 5 | 97,75% | 71,90% | 82,86% |
| | | 0,75 | 26 | 4 | 100,00% | 69,49% | 82,00% |
| | | 0,80 | 40 | 6 | 100,00% | 60,50% | 75,39% |
| E | 18 | 0,70 | 18 | 4 | 100,00% | 17,47% | 29,74% |
| | | 0,75 | 26 | 3 | 100,00% | 12,65% | 22,46% |
| | | 0,80 | 35 | 2 | 100,00% | 7,23% | 13,48% |

### 5.2.3 Agrupaments Mean-Shift

Amb els retalls de les imatges, on es troben els possibles rostres, generats en els experiments anteriors de Viola & Jones s'han realitzat les proves en aquest apartat, ja que les imatges de cada grup seran agrupades segons els seus vectors característics obtinguts de la llibreria OpenFace [20].

Els paràmetres utilitzats per les proves fetes amb els agrupaments seran: Factor 1.2, min neighbors 5, min size 5% i maxSize 35%, amb els filtres haardcascade_frontal_alt2.xml i haardcascade_profileface.xml.

#### 5.2.3.1 Agrupament general

Per analitzar l'eficàcia d'aquest sistema d'agrupament, es realitza el següent experiment, en el qual s'agruparan tots els grups per a comprovar el grau de similitud de les cares en un mateix clúster si són de la mateixa persona.

Cal esmentar que aquest sistema d'agrupament no fa servir mètodes concrets com passava en el jeràrquic, sinó que té uns valors de referència que es pot estimar tal com s'explica en l'apartat 4.3.1.

Com mostra a la taula 15, els resultats no són molt prometedors, ja que si es comparen amb els resultats obtinguts en les taules 7, 8, 9, 10 i 11 de l'agrupament jeràrquic, els resultats són millors.

Taula 15: Agrupament MeanShift

| Grup | Cares Detectades | Errors | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 0 | 1 | 1,52% | 1,73% | 1,62% |
| B | 124 | 40 | 12 | 2 | 86,48% | 43,97% | 58,08% |
| C | 181 | 75 | 0 | 1 | 13,48% | 5,61% | 7,92% |
| D | 135 | 15 | 8 | 3 | 37,00% | 57,81% | 45,12% |
| E | 65 | 18 | 5 | 1 | 35,00% | 9,86% | 15,38% |

#### 5.2.3.2 Agrupament amb Pearson

Per a millorar l'agrupació d'aquest agrupament, s'ha plantejat comprovar la consistència dels clústers que genera per mitjà del coeficient de Pearson, ja que si troba elements que no tinguin un mínim de similitud, aquests seran tret del clúster. Les imatges de cada clúster seran revisades entre totes elles i si la comparativa no arriba a un valor mínim, la imatge serà descartada. En aquest experiment, es comprovaran tots els grups, on s'aplicaràn els valors mínims de 0,70 i 0,80 del coeficient de Pearson.

Taula 16: Agrupament MeanShift aplicant Pearson

| Grups | Possibles Cares | Errors | Pearson | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 0.70 | 479 | 0 | 0% | 0% | 0% |
| | | | 0.80 | 479 | 0 | 0% | 0% | 0% |
| B | 124 | 40 | 0.70 | 10 | 4 | 88,45% | 44,35% | 59,08% |
| | | | 0.80 | 37 | 3 | 100,00% | 34,59% | 51,40% |
| C | 181 | 75 | 0.70 | 106 | 0 | 0% | 0% | 0% |
| | | | 0.80 | 106 | 0 | 0% | 0% | 0% |
| D | 135 | 15 | 0.70 | 81 | 2 | 100,00% | 33,03% | 49,65% |
| | | | 0.80 | 103 | 2 | 100,00% | 13,64% | 24,00% |
| E | 65 | 18 | 0.70 | 47 | 0 | 0% | 0% | 0% |
| | | | 0.80 | 47 | 0 | 0% | 0% | 0% |

Com es pot observar en la taula 16, al realitzar la comprovació dels clústers per mitjà del coeficient de Pearson, fa augmentar el número de possibles cares descartades, pel que redueix el Recall. Però s'ha d'observar que en les agrupacions que s'han format clústers, la seva agrupació ha estat de més del 80% inclús arribant al 100,00%.

En contra partida, de 5 grups d'imatges 3 han tingut totes les seves imatges descartades i sols s'han organitzat 2 grups, amb un elevat percentatge de precisió, però amb aquests resultats el sistema de Meanshift s'hauria de descartar.

### 5.2.4 Agrupaments a través de Spectral Clustering

Amb els retalls de les imatges, on es troben els possibles rostres, generats en els experiments anteriors de Viola & Jones, s'han realitzat les proves en aquest apartat, ja que les imatges de cada grup seran agrupades, amb el mètode del **Spectral clustering**, 4.3.1, segons els seus vectors característics obtinguts de la llibreria OpenFace [20]. Els paràmetres utilitzats per les proves fetes amb els agrupaments seran: Factor 1.2, min neighbors 5, min size 5% i maxSize 35%, amb els filtres haardcascade_frontal_alt2.xml i haardcascade_profileface.xml.

Aquest tipus d'agrupament disposa de diversos paràmetres per ser configurat, aquests paràmetres són " eigen_solver " , " affinity " i " n_init ". Amb el tipus de dades que s'utilitza en aquest treball, la configuració més optima que s'ha trobat és " eigen_solver = None " , " affinity = 'nearest_neighbors' " i " n_init = 500 ".

#### 5.2.4.1 Agrupament general

En aquest experiment, s'aplicarà l'agrupament Spectral clustering en els grups de testeig, del A al E. Es comprovarà l'eficàcia d'agrupar imatges tenint com a suposat resultat que cada clúster estigui format per les imatges d'una persona en concret, i que es generin tants clústers com persones detectades.

Taula 17: Agrupament Spectral clustering

| Grups | Possibles Cares | Errors | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 0 | 4 | 47,01% | 38,34% | 42,23% |
| B | 124 | 40 | 0 | 4 | 91,43% | 44,76% | 60,09% |
| C | 181 | 75 | 0 | 6 | 45,00% | 18,22% | 25,94% |
| D | 135 | 15 | 0 | 4 | 70,75% | 75,00% | 72,82% |
| E | 65 | 18 | 0 | 5 | 95,12% | 22,94% | 36,97% |

Com es pot observar en la taula 17, els valors obtinguts per aquest agrupament són molt millor que l'agrupament MeanShift. També, cal destacar que aquest agrupament no ha generat clústers únics, com s'observa en la columna de descarts.

#### 5.2.4.2 Agrupament amb Pearson

Per a millorar l'agrupació abans comentada, s'ha plantejat l'opció de comprovar la consistència dels clústers. Aquesta consistència es comprovaria per mitjà del coeficient de correlació de Pearson ( Descrit a 3.3), ja que es tractaria de comparar els elements entre ells per mitjà de Pearson i poder obtenir un valor de similitud. El coeficient de correlació de Pearson és una mesura normalitzada entre -1 i 1, on -1 representa que no hi ha simúlid entre elements i 1 representa que els elements són idèntics. Gràcies a la comparativa, es vol extreure dels clústers els elements que no obtinguin un valor mínim. Les imatges extretes es consideraran descartades.

En aquest experiment es comprovaran tots els grups, on s'aplicarà els valors mínims del coeficient de Pearson de 0,70 i 0,80.

Taula 18: Agrupament Spectral clustering aplicant el coeficient de Pearson

| Grup | Cares Detectades | Errors | Pearson | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 0.70 | 349 | 2 | 87,61% | 16,96% | 28,41% |
| | | | 0.80 | 361 | 5 | 95,15% | 15,75% | 27,03% |
| B | 124 | 40 | 0.70 | 11 | 3 | 98,44% | 44,68% | 61,46% |
| | | | 0.80 | 21 | 3 | 100,00% | 43,70% | 60,82% |
| C | 181 | 75 | 0.70 | 69 | 6 | 83,33% | 10,07% | 17,97% |
| | | | 0.80 | 103 | 1 | 100,00% | 1,05% | 2,08% |
| D | 135 | 15 | 0.70 | 31 | 4 | 98,98% | 59,00% | 73,93% |
| | | | 0.80 | 50 | 3 | 98,63% | 41,41% | 58,33% |
| E | 65 | 18 | 0.70 | 17 | 4 | 100,00% | 18,07% | 30,61% |
| | | | 0.80 | 38 | 2 | 100,00% | 5,42% | 10,29% |

Com es pot observar en la taula 18, a l'utilitzar la referència del coeficient de Pearson la precisió de l'agrupament augmenta de manera notòria, ja que el valor més baix és de 83,33%. Per contra partida, al comprovar la similitud dels clústers amb el coeficient de Pearson ha donat com a resultat un increment de les possibles cares descartades, fent reduir el valor de Recall.

### 5.2.5 Discriminador de Roba - Extractor

En aquesta secció, es mostren els resultats obtinguts a la fase d'experimentació amb els histogrames de color dels píxels corresponents a roba. Primer mostrem els resultats quan implantem i comparen els diferents espais de color per a diferenciar entre diverses situacions.

**a) Comparativa del diferents espais de color**

En els experiments que es presenten a continuació, es comprovarà la utilitat del vector característic de la roba aplicant-lo amb diferents espais de color, incorporant-lo al vector característic dels rostres.

Per analitzar els diferents espais de color, aquesta s'han contrastat amb 4 tipus de situacions ja definides. Aquestes situacions són:

- Roba = Persona = : Refereix que en les imatges, la roba i la persona que apareixen són iguals, es a dir, es la mateixa roba i la mateixa persona, pel que serà la referència per detectar igualtat de roba.

- Roba != Persona = : Refereix que en les imatges, la roba és diferent, però la persona és la mateixa, pel que tenim la referència per discriminar/descartar la igualtat de roba.

- Roba != Persona != : Refereix que en les imatges, la roba i la persona són completament diferents, pel que tenim la referència per discriminar/descartar la igualtat de roba.

- Roba ≈ Persona != : Refereix que en les imatges, la roba no és igual, però és molt similar i les persones són diferents, pel que aquí tenim la referència del valor intermedi i complicat per comprovar la discriminació o no de la igualtat de roba.

En la taula 19, es mostren els resultats de les comparatives entre vectors. S'han abreviat paraules on "Bhatta" fa referencia a Bhattacharyya i "Pear"fa referencia al Coeficient de Pearson. En aquesta taula es mostren els valors en calcular les distàncies de Bhattacharyya i Pearson entre els vectors característics de la roba, compostos per 24 bits. També mostren el càlcul de les distàncies si els vectors s'han obtingut excloent pell o no.

Taula 19: Comparativa entre imatges de Roba i espais de color

| | | YCrCb | | HSV | | BGR | | Gray | |
|---|---|---|---|---|---|---|---|---|---|
| Tipus Imatge | Pell | Bhatta. | Pear. | Bhatta. | Pear. | Bhatta. | Pear. | Bhatta. | Pear. |
| Roba = | Sense | 0,448 | 0,455 | 0,144 | 0,443 | 0,126 | 0,454 | 0,091 | 0,120 |
| Persona = | Amb | 0,060 | 0,434 | 0,179 | 0,334 | 0,183 | 0,064 | 0,191 | 0,101 |
| Roba != | Sense | 0,161 | 0,582 | 0,457 | 0,110 | 0,665 | -0,089 | 0,734 | -0,296 |
| Persona = | Amb | 0,139 | 0,350 | 0,487 | 0,022 | 0,679 | -0,287 | 0,722 | -0,064 |
| Roba != | Sense | 0,177 | 0,582 | 0,344 | 0,398 | 0,437 | 0,409 | 0,635 | 0,018 |
| Persona != | Amb | 0,185 | 0,586 | 0,378 | 0,233 | 0,470 | 0,344 | 0,595 | 0,178 |
| Roba ≈ | Sense | 0,144 | 0,856 | 0,251 | 0,676 | 0,408 | 0,203 | 0,493 | 0,118 |
| Persona != | Amb | 0,149 | 0,854 | 0,255 | 0,603 | 0,427 | 0,160 | 0,465 | 0,215 |

Taula 20: Comparativa de roba entre espais de color i distancia euclidiana

| Tipus Imatge | Pell | YCrCb Eucliana | HSV Eucliana | BGR Eucliana | Gray Eucliana |
|---|---|---|---|---|---|
| Roba = | Sense | 38.577.066 | 13.084.382 | 5.029.236 | 1.710.908 |
| Persona = | Amb | 40.440.444 | 14.747.574 | 5.193.730 | 1.885.364 |
| Roba != | Sense | 69.105.144 | 31.109.258 | 43.305.828 | 15.971.394 |
| Persona = | Amb | 72.778.868 | 35.120.114 | 42.161.112 | 15.593.070 |
| Roba != | Sense | 50.560.919 | 17.191.315 | 29.662.969 | 13.400.267 |
| Persona != | Amb | 60.522.309 | 19.500.343 | 35.128.727 | 14.497.893 |
| Roba ≈ | Sense | 24.497.566 | 16.188.002 | 55.433.932 | 22.091.666 |
| Persona != | Amb | 23.023.087 | 15.731.123 | 51.925.945 | 20.628.433 |

En la taula 19, es pot comprovar que el coeficient de Pearson s'aproxima al resultat esperat, però en alguns casos no és fiable, ja que dóna similitud, quan la roba no és igual o similar.

A la taula 20 es mostra el càlcul de la distància euclidiana entre els vectors característic de la roba, on el caràcter del '.' representa a les centenes. Es pot observar que la mesura depèn de la composició de la imatge i els valors que dóna són molt elevats.

Els valors obtinguts pel Coeficient de Pearson i la distància de Bhattacharyya són més òptimes per fer comparativa gràcies a la seva normalització, així es pot concretar millor la similitud de les imatges.

Els rangs de valors són entre -1 a 1 per Pearson, i 0 a 1 per Bhattacharya, on a l'obtenir un resultat pròxim al 1, vol dir, màxima similitud amb el coeficient de Pearson, mentre que tenir un resultat pròxim a 0, vol dir, màxima similitud amb la distància de Bhattacharyya.

Taula 21: Comprovar distancies amb els Canals H i V de l'espai HSV

| | | HSV Canals HV | |
|---|---|---|---|
| Tipus Imatge | Pell | Bhatta. | Pear. |
| Roba = | Sense | 0,180 | 0,349 |
| Persona = | Amb | 0,223 | 0,657 |
| Roba != | Sense | 0,494 | 0,128 |
| Persona = | Amb | 0,463 | 0,100 |
| Roba != | Sense | 0,420 | 0,396 |
| Persona != | Amb | 0,457 | 0,259 |
| Roba ≈ | Sense | 0,221 | 0,724 |
| Persona != | Amb | 0,223 | 0,657 |

També val esmentar que l'espai de colors HSV segons la taula 19 representa l'opció més optima pels valors obtinguts i la similitud amb els valors comparats, juntament amb la mesura de Bhattacharyya.

Per comprovar millor els valors obtinguts amb HSV es comproven altres configuracions, com eliminar la lluminositat de l'espai de color. En l'espai HSV el canal S representa la saturació i això afecta la llum, per això s'han comprovat els valors utilitzant sols els canals H i V, on el resultat és mostra a la taula 21.

Com es pot observar a la taula 21, els valors s'han modificat, però continuen estant distribuïts de forma similar, excepte la comparació entre vectors de la mateixa roba, on el valor de la distancia Bhattacharyya és bastant inferior en l'apartat sense pell que amb pell.

**b) Concatenació dels vectors característics de roba i cares**

Per comprovar la utilitat del vector característic de la roba, es proposen dos tipus d'experiments. En el primer es proposa concatenar el vector característic de la roba amb el vector característic de rostre i normalitzar-los. Una vegada creat el vector característic resultant de la combinació, aquest s'utilitzaria en els agrupaments provats anteriorment per comprovar si es classifiquen millor les imatges. En el segon es proposa comparar els vectors característics de la roba entre ells per comprovar si amb la distància de Bhattacharyya es pot concretar si les persones de les imatges porten la mateixa roba o similar.

Per comprovar l'eficàcia de combinar els vectors característics del rostre i la roba, s'han concatenat i normalitzat, creant un sol vector característic de 132 bits. Amb el vector resultant es comprova si els diferents sistemes d'agrupament tenen millora a l'hora de classificar les imatges.

En la taula 22 es pot observar els valors de l'agrupació jeràrquica amb el mètode average. Si es compara amb les taules de l'apartat 5.2.1, es pot observar que donen millor resultat el vector característic de cares.

Taula 22: Agrupament jeràrquic amb vector normalitzat de roba i cara

| Grup | Cares Detectades | Errors | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 4 | 15 | 29,95% | 37,43% | 33,27% |
| B | 124 | 40 | 8 | 4 | 86,84% | 11,19% | 19,82% |
| C | 181 | 75 | 12 | 11 | 50,54% | 0,35% | 0,59% |
| D | 135 | 15 | 5 | 9 | 83,48% | 38,39% | 52,59% |
| E | 65 | 18 | 0 | 5 | 55,32% | 5,49% | 9,99% |

Taula 23: Agrupament MeanShift amb vector normalitzat de roba i cara

| Grup | Cares Detectades | Errors | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 12 | 4 | 2,47% | 2,37% | 2,42% |
| B | 124 | 40 | 17 | 2 | 89,55% | 41,04% | 56,28% |
| C | 181 | 75 | 10 | 1 | 1,22% | 0,48% | 0,69% |
| D | 135 | 15 | 17 | 2 | 58,42% | 54,17% | 56,21% |
| E | 65 | 18 | 7 | 1 | 37,14% | 8,72% | 14,13% |

En la taula 23 es poden observar els valors de l'agrupació per mitjà de MeanShift. Es pot observar que el resultat, en general, és menys òptim utilitzar el vector característic normalitzat del rostre i roba.

Taula 24: Agrupament Spectral clustering amb vector normalitzat de roba i cara

| Grup | Cares Detectades | Errors | Descarts | Clusters | Precision | Recall | F-Score |
|---|---|---|---|---|---|---|---|
| A | 533 | 54 | 0 | 1 | 16,70% | 27,02% | 20,64% |
| B | 124 | 40 | 0 | 4 | 72,62% | 44,25% | 54,99% |
| C | 181 | 75 | 0 | 3 | 24,66% | 7,56% | 11,58% |
| D | 135 | 15 | 0 | 5 | 70,45% | 31,63% | 43,66% |
| E | 65 | 18 | 0 | 4 | 83,33% | 6,1% | 11,36% |

En la taula 24 es poden observar els valors de l'agrupació per mitjà del Spectral clustering. Es pot observar que el resultat, en general, és menys òptim utilitzar el vector característic normalitzat del rostre i roba.

**c) Comparativa dels vectors característics de roba**

Es vol comprovar l'eficàcia de comparar diversos vectors característics de roba. S'utilitzarà un conjunt format per 22 imatges, les quals tenen característiques similars i diferents entre elles. Amb aquestes comparacions es mesurarà la precisió que té utilitzar la distància de Bhattacharyya per discriminar la roba si és diferent.

S'utilitzarà l'espai de color HSV, ja que segons la taula 19 és el resultat més òptim i s'utilitzarà un vector característic de dos canals el H i el V, per eliminar la lluminositat. Com mostra la taula 21, els resultats de l'espai de color amb els canals H i V amb l'exclusió de la pell donen uns resultats molt òptims. Per aquest

motiu s'ha volgut comprovar l'eficàcia d'eliminar la llum i utilitzar un vector de 16 bits.

Taula 25: Comparativa entre imatges

| Comparatives | Tall Bhatta | Pell | Presicio |
|---|---|---|---|
| 190 | 0,25 | Sense | 85,26% |
| | | Amb | 90,00% |
| | 0,35 | Sense | 60,00% |
| | | Amb | 71,58% |

Els resultats que es mostren a la taula 25, l'últim camp indica la precisió; aquest punt indica el percentatge d'encerts en discriminar si les imatges comparades tenien o no la mateixa roba. Noteu que en cap moment s'ha comparat la mateixa imatge entre si.

Com es pot observar a la taula 25 quan es calcula la distància de Bhattacharyya sense la pell, la precisió augmenta en tots els casos explorats.

# 6 Conclusions i treball futur

En aquest treball, s'ha proposat la re-identificació de persones per mitja de l'agrupació de dades recopilades. Aquestes persones son les que apareixen en imatges gravades per dispositius portables, amb la finalitat de poder crear un àlbum d'interaccions socials, a través del reconeixement. Per la re-identificació es proposa agrupar els vectors descriptius dels rostres detectats en les imatges egocèntriques, afegint informació contextual en els vectors, concretament la roba. A més, s'implementa la distorsió de rostres obtinguts per mitjà d'imatges egocèntriques, perquè la persona que apareix no sigui reconeguda.

En el cas de la distorsió dels rostres, es pot concloure, després dels experiments exposats en l'apartat 5.1.4, que l'eficàcia de la distorsió dels rostres està lligada a l'eficàcia de l'algorisme Viola & Jones, ja que la distorsió de les cares no té complicació, on l'usuari pot concretar el grau de distorsió. La complicació està en la localització del rostre, on l'algorisme Viola & Jones té uns resultats relativament òptims, però s'han de tenir presents els paràmetres que el configuren. Com es pot apreciar en la taula 6, si es volen localitzar totes les cares possibles, s'ha de definir el paràmetre min_neighbores amb un valor molt baix, però això genera un increment de falsos positius i obtenir cares de menor qualitat, pel que seran descartades pels agrupaments posteriors i augmentaran els falsos negatius. Així que l'ús de l'algorisme Viola & Jones s'ha de concretar segons el tipus d'imatges que es volen processar i si es volen recopilar cares amb una certa definició. També ha estat molt relevant l'ús dels paràmetres d'aquesta funció per tal d'adaptar-la al nostre problema.

A més, s'han implementat diversos algoritmes d'agrupament. Els agrupaments Jeràrquic i Spectral clustering són més òptims. S'ha pogut comprovar que el mètode de l'enllaçament *average* de l'agrupament jeràrquic és el més efectiu. Es pot concloure que pel tipus de dades que es processen en aquest treball, l'agrupament que realitza MeanShift no és òptim, pel que s'hauria de descartar. Els dos tipus d'agrupament, jeràrquic i Spectral clustering, donaven uns resultats de precisió poc estables, amb valors que oscil·laven entre el 45% i el 95%, però en revisar la consistència dels clústers amb el coeficient de Correlació de Pearson, aquest marge ha millorat estant entre el 80% i 100%. Pel que es pot concloure que pel tipus de dades que s'analitzen, el coeficient de Correlació de Pearson comprova molt bé la similitud entre els rostres. Per contra partida, utilitzar el Coeficient de Correlació de Pearson per comprovar la consistència dels clústers crea més falsos negatius al descartar rostres, perquè no quedin mal classificats. Això provoca la reducció del valor recall.

S'ha inclòs informació contextual per a la classificació, a més s'han implementat diferents distàncies per a mesurar l'agrupament. La comparativa de la roba per mitjà de la distància de Bhattacharyya ha donat uns resultats molt òptims. S'han comparat vectors formats pels canals H i V de l'espai de colors HSV, arribant a una precisió del 90% en comprovar si les robes de vàries imatges eren iguals o similars. Els valors extrets de la roba han estat excloent els píxels de la pell.

En aquest treball, s'ha començat a tractar la possibilitat de millorar l'agrupació

de persones incloent informació contextual de la roba que porta la persona detectada. Els resultats no han estat òptim, com es pot observar en les taules de l'apartat 5.2.5.

Com a linea de treball futura, per un costat, es seguirà experimentant amb altres característiques descriptives de la roba, com per exemple textures, o costures, entre altres possibilitats. Per un altre costat, s'utilitzaran les *convolutional neural networks* per a l'extracció de característiques i classificació d'aquestes. A més, ja que s'ha demostrat que l'eficàcia de la mesura de bhattacharyya en la classificació de la roba similar és molt òptima. Seria interessant implementar aquesta distància en el sistema d'agrupament.

# Referències


[1] Gemmell, J.; Lueder, R.; Bell, G.: *The MyLifeBits Lifetime Store*, Microsoft Research,
https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/etp2003.pdf , Novembre 2003.

[2] Gabriel Oliveira-Barra and Marc Bolaños and Estefanía Talavera and Adrián Dueñas and Olga Gelonch and Maite Garolera: *Serious Games Application for Memory Training Using Egocentric Images*, Novembre 2017.

[3] Zheng, W.; Li, X.; Liao, S; Lai, J.; Gong, S.: *Partial Person Re-identification*, Sun Yat-sen University, China,
https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Zheng_Partial_Person_Re-Identification_ICCV_2015_paper.pdf , Novembre 2015.

[4] Zheng, L.; Yang, Y.; Hauptmann, A.: *Person Re-identification: Past, Present and Future*, Sun Yat-sen University, China,
https://arxiv.org/pdf/1610.02984v1.pdf , Octubre 2015.

[5] Shi, X; Chen, Z; Wang,H, Yeung, D.: *Convolutional LSTM Network: A Machine LearningApproach for Precipitation Nowcasting*
https://arxiv.org/pdf/1506.04214.pdf, Department of Computer Science and Engineering, Hong Kong, University of Science and Technology, 2015.

[6] Umberson, D; Karas Montez, J.: *Social Relationships and Health: A Flashpoint for Health Policy*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3150158/, Journal of health and social behavior, vol 51, no. 1 suppl, pp S54-S66, 2010

[7] Aghaei, M.; Dimiccoli, M.; Radeva, P.: *With Whom Do I Interact? Detecting Social Interacti-ons in Egocentric Photo-streams*, Universitat de Barcelona,
https://arxiv.org/pdf/1605.04129v1.pdf , Maig 2016.

[8] Aghaei, M.; Dimiccoli, M.; Canton Ferrer, C ; Radeva, P.: *Social Style Characterization from Egocentric Photo-streams*, Universitat de Barcelona,
https://arxiv.org/pdf/1709.05775.pdf , Setembre 2017.

[9] Sun, Q.; Schiele, B.; Fritz, M.: *A Domain Based Approach to Social Relation Recognition*, Max Planck Institute for Informatics, Saarland Informatics Campus,
https://arxiv.org/pdf/1704.06456.pdf , Abril 2017.

[10] Lin, Y.; Zheng, L.; Wu, Y.; Yang, Y.: *Improving Person Re-identification by Attribute and Identity Learning*, University of Technology Sydney,
https://arxiv.org/pdf/1703.07220.pdf , Abril 2017.

[11] Alameda-Pineda, X.; Staiano, J.; Subramanian, R.; Batrinca, L.; Ricci, E.; Lepri, B.; Lanz, O.; Sebe, N.: *Salsa: A novel dataset for multimodal group behaviour analysis*, 2015

[12] Vinciarelli, A.; Pantic, M.; Bourlard, H.: *Social signal processing: Survey of an emerging domain*, Image and Vision Computing, vol. 27, no. 12, pp. 1743–1759, 2009

[13] Kendon, A.: *Conducting interaction: Patterns of behavior in focused encounters*, CUP Archive vol. 7, 1990

[14] Aghaei, M.; Domiccoli, M.; Radeva, P.: *Multi-face tracking by extended bag-of-tracklets in egocentric photo-streams*, Journal of Computer Vision and Image Understanding. doi:10.1016/j.cviu.2016.02.013. arXiv preprint arXiv:1507.04576, 2016

[15] Xiong, Y.; Quek, F.: *Meeting room configuration and multiple camera calibration in meeting analysis*, In Proceedings of the 7th international conference on Multimodal interfaces, pages 37–44. ACM `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.424&rep=rep1&type=pdf`, 2005.

[16] Amato, G.; Debole, F.; Falchi, F.; Gennaro, C.; Rabitti, F.: *Large scale indexing and searching deep convolutional neural network features*, In International Conference on Big Data Analytics and Knowledge Discovery, pages 213–224. Springer, 2016.

[17] Open Source Computer Vision Librarym: *Documentació OpenCV*, `https://docs.opencv.org/2.4.11/index.html` , 2015.

[18] Viola, P; Jones, M.: *Robust Real-Time Face Detection*, `http://www.ipol.im/pub/art/2014/104/article.pdf`, International Journal of Computer Vision 57(2), 137–154, 2004.

[19] Aghaei, M; Dimicooli, M; Radeva, P.: *ALL THE PEOPLE AROUND ME: FACE DISCOVERY IN EGOCENTRIC PHOTO-STREAMS*, `https://arxiv.org/pdf/1703.01790.pdf`, University of Barcelona, Computer Vision Center, 2017.

[20] www.github.com/cmusatyalab : *OpenFaces* `http://cmusatyalab.github.io/openface/` , 2016.

[21] www.sciPy.org : *Hierarchical clustering* `https://docs.scipy.org/doc/scipy-0.18.1/reference/cluster.hierarchy.html#module-scipy.cluster.hierarchy` , 2016.

[22] scikit learn : *sklearn.cluster.MeanShift* `http://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html` , 2017.

[23] Gomez Villegas, M.: *Karl Pearson, el Creador de la Estadística Matemática* `http://www.mat.ucm.es/~villegas/ArtPearson2007.pdf`, Dpto. de Estadística e Investigación Operativa, Fac. de CC Matemáticas, Universidad Complutense de Madrid , 2007.

[24] Marín Reyes, P. A.: *Estudio comparativo de medidas de distancia para histogramas en problemas de reidentificación*, Tesis de Màster per a la Universitat de las Palmàs de Gran Canària, 2015

[25] Funes A.: *Agrupamiento Conceptual Jerárquico Basado en Distancias*, Tesis de Master per a la Universitat Politècnica de València, 2008

[26] Alegre Gutierrez, E.; Pajares Martinsanz, G.; de la Escalera Hueso, A.: *Conceptos y Métodos en Visión por Computador*, ISBN: 978-84-608-8933-5, Espanya, Juny 2016.

[27] Pignataro N.; Figueredo, G.: *Spectral Clustering*, 2008.

[28] Davis T.A; Altevogt R.: *Golden mean of the human body*, 1979

[29] Cola, A.: *GitHub: Face Detection (Codi TFG)*, Universitat de Barcelona, `https://github.com/ColtronRuso/FD2017TFG`, 2018.