

Reconeixement facial a través de vídeo.

Lucas Gonzalo Priego, Jakob Kotnik Lopez

1 INTRODUCCIÓ

La proposta del nostre projecte es basa en el reconeixement facial. Concretament aprofitarem el nostre projecte de Robòtica per relacionar-ho amb l'assignatura de VC. Per tal de situar en context el nostre projecte de visió per computador, cal explicar el funcionament del nostre robot. Aquest constarà d'una càmera situada sobre una plataforma giratòria, situada a l'alçada del rostre de un nombre determinat de persones que estaran assegudes al voltant d'una taula. L'objectiu d'aquesta càmera serà reconèixer la identitat de les persones que formen part d'aquesta taula, y determinar si són les desitjades. Per tant podríem resumir el nostre problema com un projecte de reconeixement facial.

A continuació hem de tenir en compte els diferents entorns en el que ens podem trobar el nostre robot y com aquest afectarà all reconeixement. Primerament, hem de saber que trobarem aspectes en comú en tots els espais, concretament sempre la nostra càmera treballarà a una alçada y distancia similar respecte els rostres a reconèixer. Això es deu a que com hem explicat es trobarà sempre de manera estàtica sobre una superfície estable com pot ser una taula. Per tant el nostre model haurà de deprecier els rostres que es trobin fora d'aquests rangs.

Per acabar, hem d'entendre que també haurà característiques ambientals que poden influir en el reconeixement. Poden ser característiques com la lluminositat, estem a contra llum? estem situats en un entorn obert o tancat?. Finalment també hem de saber que les mostres utilitzades per entrenar el model, s'han de correspondre amb les de test, es ha dir que el model s'ha d'entrenar amb unes característiques invariables d'un rostre, que si canvien no es podrà reconèixer aquest mateix rostre. Es a dir el model s'haurà d'entrenar en un context similar al que volem realitzar el test.

2 OBJECTIUS

Centrant-nos en l'objectiu del nostre projecte, es correspon en la realització d'un programa que sigui capaç de reconèixer a les persones que es troben al voltant d'una taula.

Per tal de fer-ho realitzarem un programa en python, que dividirem en diferents funcions que ens permetran aconseguir l'objectiu.

Primerament realitzarem una fase d'introducció a la detecció de rostres, per tal d'entendre com funcionarà aquesta primera part inicial del projecte. Per tant aquesta primera fase tindrà com objectiu detectar els diferents rostres que trobem en una imatge.

La segona fase correspondrà a la detecció de rostres en un vídeo. Concretament realitzarem dues parts, la primera d'elles tindrà com objectiu detectar un rostre en un vídeo donat, es ha dir d'internet, que tingui unes condicions favorables per tal de que el reconeixement sigui correcte. La segona part tindrà com ha finalitat reconèixer els rostres d'un vídeo, igual que la part ja esmentada, però en aquest cas d'un vídeo en directe utilitzant la càmera que utilitzarem per realitzar la resta del projecte.

Una vegada superada aquesta fase inicial, ens centrarem en la part de reconeixement facial. Aquesta ultima fase tindrà com objectiu realitzar tota la part necessària pel reconeixement facial. Més endavant en aquest document entrarem en detall de la realització de cadascuna d'aquestes fases

Finalment haurem de validar les nostres dades per veure que tant bo ha estat el nostre reconeixement. Aquesta fase té com ha objectiu determinar l'accuracy del nostre model, que haurem d'intentar que sigui la més elevada possible, per tal de considerar el nostre projecte robust, y poder extreure les conclusions finals.

3 ESTAT DE L'ART

El reconeixement facial consisteix en el desenvolupament de tècniques aplicades en què un programa informàtic és capaç de reconèixer automàticament una cara, associant-ho amb una identitat determinada. Això és possible mitjançant una anàlisi de les característiques facials d'un subjecte extretes d'una imatge o d'un fotograma clau d'una font de vídeo, i la comparació amb una base de dades preexistent. Actualment s'utilitzen diferents tècniques i algorismes per abordar aquest tipus de problema i els més utilitzats serien:

- Tècniques de visió tridimensional
- Reconeixement de patrons
- Proves one-to-one
- Algorismes de detecció de punts clau de rostre com:
 1. MTCNN
 2. AdaBoost
 3. SVM
 4. ConvNets

Una vegada esmentades quines són les propostes en l'actualitat, ens centrarem en quines utilitzarem nosaltres, per que les escollim y en que consten. Primerament cal parlar de la detecció de rostres.

El nostre projecte es basarà en un document publicat per Paula Viola y Michael Jones l'any 2001 a la universitat de Cambridge. En aquest document que te com a títol "Rapid Object Detection using a Boosted Cascade of Simple Features", on es parla d'una perspectiva del aprenentatge automàtic per la detecció visual d'objectes d'una manera extremadament rapida y aconseguint unes altes tasses de reconeixement. Aquest document es distingeix per tenir tres contribucions clau.

La primera d'elles fa referencia a la representació de la imatge anomenada Imatge Integral, la segona tracta sobre l'aprenentatge, concretament d'un algorisme basat en Adaboost, que selecciona una petita quantitat de característiques visuals

crítiques d'un conjunt més gran y produeix classificadors molt eficients. Finalment, la tercera d'elles fa referència a un mètode per combinar classificadors cada vegada més complexos, per tal de formar una cascada de classificadors que permet que el seu treball de processament y descartat d'imatge sigui el més eficient possible.

Per tant en el nostre cas utilitzarem un d'aquests classificadors "cascade" especialitzat en la detecció de cares. Un aspecte que ens sembla interessant esmentar és quines característiques utilitza el mètode esmentat anteriorment per detectar objectes. Ho fa de la següent manera:

El procediment de detecció d'objectes els classifica tenint en compte el valor de les característiques simples

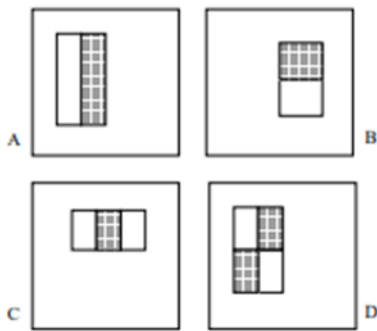


Figura 1: Exemple de característiques de rectangle mostrades en relació amb la finestra de detecció envoltant. La suma dels píxels que es troben dins dels rectangles blancs es resten de la suma de píxels als rectangles grisos. Característiques de dos rectangles (A) i (B). La figura (C) mostra un rectangle de tres característiques, i (D) mostra un rectangle de quatre característiques.

Concretament aquestes característiques s'anomenen característiques de Haar. En la bibliografia aportem més informació sobre aquest apartat.

Pel que fa el reconeixement facial utilitzarem dos mètodes per crear dos models diferents, el primer d'ells basat en EigenFaces i el segon basat en Fisher faces. El primer d'aquest es basa en les característiques invariables d'una sèrie de rostres, o el que es el mateix, els vectors propis de la matriu de covariància del conjunt d'imatges que formen la base de dades. A més aquest és un mètode que utilitza una tècnica per tal de reduir la dimensionalitat del problema, anomenada PCA (Principal component anàlisi). Les característiques principals d'aquesta tècnica són: PCA revela l'estructura interna de les dades d'una manera que millor expressa la variància de les dades, ordena les diferents components per importància: podem descartar les que no aporten massa informació (reducció de la dimensió), es pot visualitzar/explicar com una translació (mean centering) seguida d'una rotació que alinea el núvol de dades als eixos principals, finalment una de les formes d'obtenir un PCA és mitjançant la descomposició en valors singulars (SVD).



Fig2 En aquesta figura podem veure un exemple de la utilització d'eigenfaces per diferents rostres.

El segon d'ells en canvi, anomenat FisherFaces, és un mètode que te en compte la forma en la que reflexa la llum i les expressions facials dels Rostres. Concretament aquest algoritme maximitza la relació entre la distribució de les classes i la distribució intra-class. També cal esmentar que aquest mètode també utilitza un mètode per reduir la dimensionalitat de les dades, concretament utilitza el mètode LDA (Linear Discriminant Analysis), o també anomenat Fisher's linear discriminant analysis. Les característiques principals d'aquesta tècnica són: Que te en compte les etiquetes de les classes en la projecció, el seu objectiu es maximitzar la separació entre classes en comptes de preservar la variància de les dades.



Fig3 En aquesta figura podem veure un exemple de la utilització de Fisher Faces per diferents rostres.

4 PROPOSTA

Una vegada descrit l'estat de l'art, parlarem de la nostra proposta, en les tres fases esmentades en l'apartat d'objectius.

La fase de reconeixement facial mantindrà la següent estructura:



Fig4: Esquema de l'obtenció del classificador

En el nostre cas no realitzarem aquest procés de forma manual, simplement utilitzarem un classificador ja proporcionat per la llibreria opencv, que ha estat creat de la manera esmentada.

Aquest classificador és "haarcascade_frontalface_default.xml". Per tant com hem dit utilitzarem aquest procés per les dues primeres fases del projecte. En quant a la implementació d'aquesta part realitzem dues funcions, anomenades "detecció img" i "deteccióVideo", les quals com el seu nom indiquen, tindran com a funció detectar els rostres que trobem en els dos casos. Si entrem al detall de la codificació d'aquestes funcions podem veure la utilització d'una funció per part del classificador, anomenada detectMultiScale, realment aquesta funció és on es realitza la detecció de rostres donats una sèrie de paràmetres que cal explicar.

El primer d'ells correspon a la imatge que volem processar, en escala de grisos, en el cas dels vídeos farem el mateix procés per cada frame.

El segon fa referència a l' scaleFactor, on especifiquem quant es reduirà la imatge, per exemple 1.2 significarà que es reduirà en un 20%, fins formar una piràmide d'imatges. Hem de tenir en compte que si aquest nombre es molt reduït el temps de processament serà molt elevat i el nombre de falsos positius es poden incrementar, en canvi si aquest numero es molt alt es pot perdre la detecció d'algun rostre.

El següent, MinNeighbors especifica el nombre de veïns que ha de tenir cada rect6angle candidat per tal de marcar-ho com positiu. En el cas de que aquest paràmetre sigui molt petit es pot donar el cas de que un mateix rostre es detecti mes d-una vegada alhora. En canvi si aquest valor es massa alt pot ser que no es detecti cap rostre.

Finalment els paràmetres "MinSize i MaxSize", s'encarreguen de delimitar els objectes que tindrem en compte, es ha dir un objecte més gran de max size o mes petit que min size no es tindran en compte.

Una vegada processada la imatge simplement dibuixem un quadrat per cadascun dels rostres detectats.

Com sabem, la següent part del nostre projecte fa referència al reconeixement facial. Per tal d'implementar-ho dividim el codi en tres funcions.

La primera d'elles "RoconocerJugadorTarin", te com objectiu tenir 300 imatges de cada rostre de manera independent, per tal de ser utilitzades posteriorment en la fase de training del nostre model. Aquesta funció utilitza els principis mencionats en la fase de detecció. Per tant per cada frame del vídeo en el que detecta un acara, realitza una captura del rostre.

La segona d'elles "Training", s'encarrega de generar dos models que seran utilitzats posteriorment per determinar la identitat de les persones que apareguin en el vídeo de test. Els dos models corresponen a EigenFaces i FisherFaces, tots dos explicats en detall a l'apartat anterior.

L'ultima "ProbarModelo" s'encarrega de realitzar el test en un vídeo pre-gravat en el que apareixen diferents rostres, tant coneguts com desconeguts per tal de determinar el bon funcionament del programa.

Podrem veure les diferent proves realitzades en el l'apartat d'experiments i resultats, on entrarem en detall sobre les dades utilitzades en cadascuna de les proves.

5 DADES I EXEMPLES

Pel que fa les dades utilitzades al projecte, hem utilitzat vídeos realitzats per nosaltres mateixos, amb la càmera que serà utilitzada en el projecte de robòtica. Aquesta es una càmera connectada al robot a través de USB.

Pel que fa el context dels vídeos hem, tingut en compte que siguin condicions que podríem afirmar com normals en les que el nostre robot es podrà trobar de manera real. Es a dir en, espais tancats, ja que ha de tenir un endoll per tal de funcionar, on els rostres es trobin a una distancia que tingui sentit, ja que com hem esmentat el robot es trobarà sobre una taula on els jugadors estaran envoltant-la. També em tingut en compte que pot haver altres rostres fora del rang esmentat, que el nostre robot haurà de ser capaç d'obviar.

Una vegada explicat com obtindrem les diferents mostres de imatge y vídeo, cal explicar la quantitat d'elles que utilitzarem posteriorment en els experiments.

Pel que fa la part inicial de detecció en una simple imatge, hem utilitzat una mostra de diferents imatges d'internet, en les que puguem trobar diferents rostres. Concretament hem utilitzat un total de 5, les quals tenen diferents condicions ambientals, el rostres es troben a diferents distancies y finalment on no tots els rostres que apareixen tenen la cara frontal respecte la càmera.

Seguidament necessitarem realitzar els diferents vídeos individuals que utilitzarem per realitzar el training. En el nostre cas hem realitzat els vídeos individuals de quatre persones. Cal tenir en compte que aquest training ha de tenir les mateixes condicions de la fase de test, si no aquesta segona fase pot donar errors. També durant la realització del vídeo hem de dir a l'usuari que realitzi diferents expressions per tal de posteriorment tenir-les en compte per l'entrenament.

Per acabar hem, gravat un conjunt de vídeos que utilitzarem com test, aquests es troben en diferents contextos, on apareixen un numero diferent de rostres, coneguts i no coneguts a diferents distancies.

6 EXPERIMENTS, RESULTATS I ANÀLISI

Dividirem els experiments en experiments de detecció i experiments de reconeixement.

Pel que fa la fase de detecció, hem realitzat dos experiments diferents, el primer d'ells per tal de testear la part de detecció de rostres en imatges, i el segon per tal de detectar els rostres en vídeo. Podem veure un exemple d'aquest experiment en la següent figura:



Figura 5: Resultat de l'execució

En la figura anterior podem veure una mostra on trobem cares a una distància molt similar però amb certes inclinacions. Com podem veure el classificador en aquest cas encerta y troba tots els rostres.

Seguidament realitzem les proves per determinar el correcte funcionament de la detecció en vídeo. Aquesta part consta de tres proves una amb un vídeo d'internet, una segona prova amb un vídeo pregravat, i finalment amb un vídeo en directe. Per tal de veure si el resultat es correcte, indiquem al programa que realitzi una captura del frame on trobem el jugador. Un exemple és el següent:

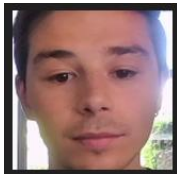


Figura 6: Captura realitzada pel programa al detectar un rostre.

Pel que fa la part de reconeixement realitzem diferent experiments, utilitzant els dos tipus de model i amb un nombre diferent de rostres implicats.

Com hem esmentat utilitzem vídeos individuals de cada participant per tal d'extreure les 300 imatges de training. El format d'aquesta part és exactament el mateix que acabem d'esmentar, però en comptes d'extreure una única imatge de cadascun dels rostres obtenim 300.

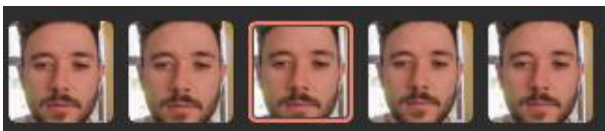


Figura 4: Seqüència d'imatges d'un participant.

Una vegada obtinguts i entrenats els nostres models realitzarem diferents proves, en diferents contextos. En el nostre cas hem realitzat proves amb dues úniques persones, en el mateix context que les dades de training, una altra prova amb quatre persones, però en aquest cas amb dos persones que s'han entrenat en aquest entorn y altres que no, finalment una prova amb quatre persones entrenades en el mateix context de la prova. En totes elles hem realitzat diverses execucions variant

els paràmetres descrits en el apartat anterior, per tal d'obtenir els millor resultats possibles en cadascun dels exemples. A continuació adjuntem captures d'algunes d'aquestes proves:

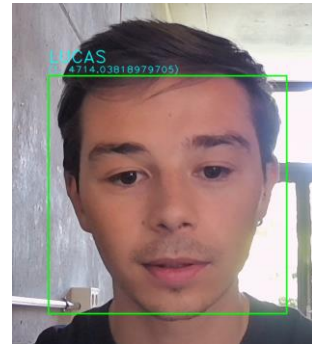


Figura 7: Captura extreta de prova real, model EigenFaces, mateix entorn, prova realitzada amb dos rostres coneguts alhora



Figura 8: Captura realitzada en la mateixa prova que la figura anterior, en aquest cas el model utilitzat correspon a Fisher Faces, en aquest cas la captura s'ha realitzat en el moment que trobem els dos rostres.

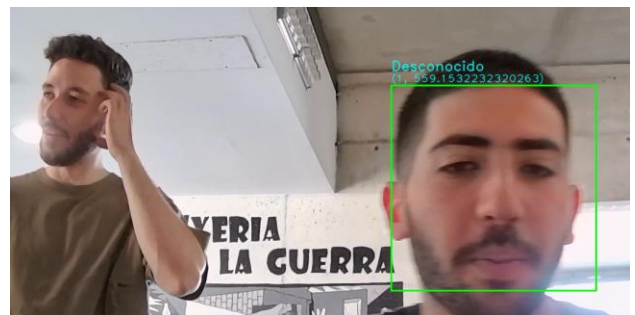


Figura 9: Captura realitzada en un altre context, amb cinc persones, com podem veure reconeix el rostre però el marca com desconegut, ja que no hem entrenat el model per que el detecti. També com podem veure obvia els rostres que es troben fora del rang de distància especificat.

Per tal de mesurar l'eficàcia del nostre model hem dividit les dades en test i training, aquestes dades que em utilitzat pel training han estat un total de 300 imatges extretes de vídeos individuals, a continuació hem eliminat aquelles que eren errònies manualment. Una vegada entrenat el nostre model, hem realitzat la fase de test, on amb 300 imatges també en el mateix context que les imatges utilitzades a training hem determinat quines d'elles eren positives, quines negatives, y quines d'elles eren o bé fals positiu o be fals negatiu.

Per les diferents proves hem obtingut una accuracy mitjana per cada model de:

Model EigenFaces: $282/300 = 94\%$

Model Fisher Faces: $293/300 = 97,6\%$

Per tant podem afirmar que el segon model es mes eficaç en el context utilitzat a les proves.

7 CONCLUSIONS

Com a conclusions del treball podem afirmar que em aconseguits els objectius plantejats al principi del projecte. Cal destacar l'eficàcia dels models realitzats, concretament els models amb les proves amb menor nombre de persones i realitzats en condicions ideals. Per condicions ideals entenem aquelles condicions ambientals que tenen relació amb el funcionament del robot i que per tant pot ser condicions reals de treball del mateix, però que tenen una il·luminació ideal. Encara que per els altres experiments obtenim pitjors resultats, cal dir que son bons, i que per tant el nostre programa s'adapta a diferents característiques ambientals, com pot ser el nombre de persones o diferents condicions de llum.

Pel que fa a coneixements, hem pogut aprofundir en aspectes relacionats amb el reconeixement, sobre tot amb tècniques com el PCA i LDA i pel que fa a aspectes no tan relacionats amb l'assignatura, em millorat els nostres coneixements de programació en python, concretament en la llibreria Opencv, i finalment hem continuat expandint els nostres coneixements de Machine learning, ja introduïts en assignatures com Aprenentatge Computacional del primer semestre.

Tots els aspectes esmentats en aquest informe seran explicats en l'exposició oral que realitzarem sobre el nostre projecte.

8 BIBLIOGRAFIA

- [1]<https://hackernoon.com/6-best-open-source-projects-for-real-time-face-recognition-vr1w34x5>. En aquest blog se'ns plantegen diferents problemes relacionats amb el reconeixement facial, i podem trobar una serie de repositoris on la gent treballa sobre aquests casos i podem trobar infinitat de variants i propostes que ens permetran reflexionar sobre el plantejament del nostre codi, concretament com estructurar el flux del programa i els aspectes a tenir en compte en el mateix.
- [2]<https://icc.fcen.uba.ar/reconocer-rostros-con-eficiencia-un-enorme-desafio-computacional/> Informació sobre l'estat de l'art.
- [3]<https://www.cienciadedatos.net/documentos/py34-reconocimiento-facial-deeplearning-python.html> Exemple de projecte