

VISIÓ PER COMPUTADOR

Short project

Ferran Conde Codorniu
Sergi Ibàñez Peña

Facultat d'Informàtica de Barcelona
Curs 2018 - 2019 : Q1

Índex de continguts

1. Objectius del projecte	2
2. Detecció d'ulls	3
Estudi estadístic	4
3. Detecció de mirada	7
4. Eina de proves	9
Resultats de l'eina de proves	14
5. Conclusions	17

1. Objectius del projecte

El problema

L'objectiu del projecte és desenvolupar un detector de mirada. Per aconseguir-ho, donada una imatge, s'ha de detectar els ulls (emprant, en aquest cas, una finestra lliscant) i posteriorment classificar si els ulls miren a càmera o no.

El conjunt de dades

Disposem de 1520 imatges de 384x286 píxels en escala de grisos. També disposem de les coordenades dels ulls per cadascuna d'aquestes imatges.

Atès que treballem amb retalls d'imatges que comprenen els dos ulls sencers (i no dos retalls d'un ull cadascun), hem de definir la mida del rectangle.

Originalment, volíem definir un rectangle per cada imatge amb mida relativa a la distància entre els ulls, i evitar així problemes quan el subjecte s'apropa o s'allunya de la càmera.

El problema és que rectangles de mides diferents generen un nombre diferent de característiques HOG i no és adequat pel classificador. Vam observar que, per exemple, amb retalls de 108x37 píxels obteníem aproximadament el doble de característiques HOG que amb retalls de 79x27 píxels.

Per tant, vam definir la mida del rectangle de retall a 80x25 píxels.

Per cada imatge, en retallem el rectangle dels ulls i n'extreiem les característiques HOG, així com els descriptors de forma. A més, atès que la proporció de mostres positives i negatives ha de ser del 5% vs 95%, per cada mostra d'ulls generem 19 retalls aleatoris de la imatge, evitant que el retall surti dels límits de la imatge i també evitant que es generi un retall prop del retall "positiu". Això ho fem per disposar d'una diferència marcada i evitar ambigüitats.

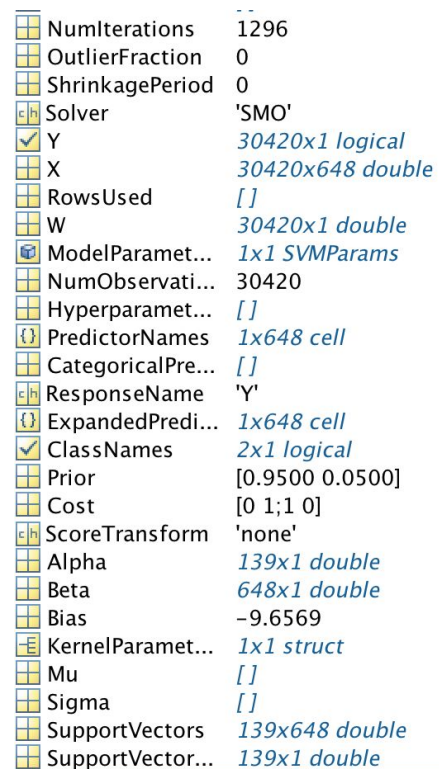
Per acabar, conforme anem generant les mostres negatives o extreient les positives, anem construint de forma paral·lela el vector d'etiquetes o *labels*, ja que es tracta d'un problema d'aprenentatge supervisat.

2. Detecció d'ulls

Havent efectuat els passos de preprocessat descrits anteriorment, podem entrenar el classificador.

El classificador emprat és un classificador de *màquina de vectors de suport* (Support Vector Machine Classifier, d'aquí en endavant SVC).

Per començar, avaluarem el classificador SVC entrenat únicament amb característiques HOG.



NumIterations	1296
OutlierFraction	0
ShrinkagePeriod	0
Solver	'SMO'
Y	30420x1 logical
X	30420x648 double
RowsUsed	[]
W	30420x1 double
ModelParamet...	1x1 SVMParams
NumObservati...	30420
Hyperparamet...	[]
PredictorNames	1x648 cell
CategoricalPre...	[]
ResponseName	'Y'
ExpandedPredi...	1x648 cell
ClassNames	2x1 logical
Prior	[0.9500 0.0500]
Cost	[0 1;1 0]
ScoreTransform	'none'
Alpha	139x1 double
Beta	648x1 double
Bias	-9.6569
KernelParamet...	1x1 struct
Mu	[]
Sigma	[]
SupportVectors	139x648 double
SupportVector...	139x1 double

Figura 2.1. Propietats del primer classificador SVC (només HOG).

Podem veure que el classificador empra només 139 de 30420 mostres com a vectors de suport. Això és una proporció del 0.457%.

Sabem que una proporció de vectors de suport acceptable per un classificador SVM és menor de 50% i, idealment, menor del 20-30%. En aquest cas, veiem que és una proporció molt baixa.

Això, a priori, pot sorprendre, ja que és sabut que els SVC no treballen gaire bé sobre conjunts de dades desequilibrats (com és el nostre cas, ja que la proporció de les dades és 5/95).

No obstant, ho podem entendre si comparem dues mostres:

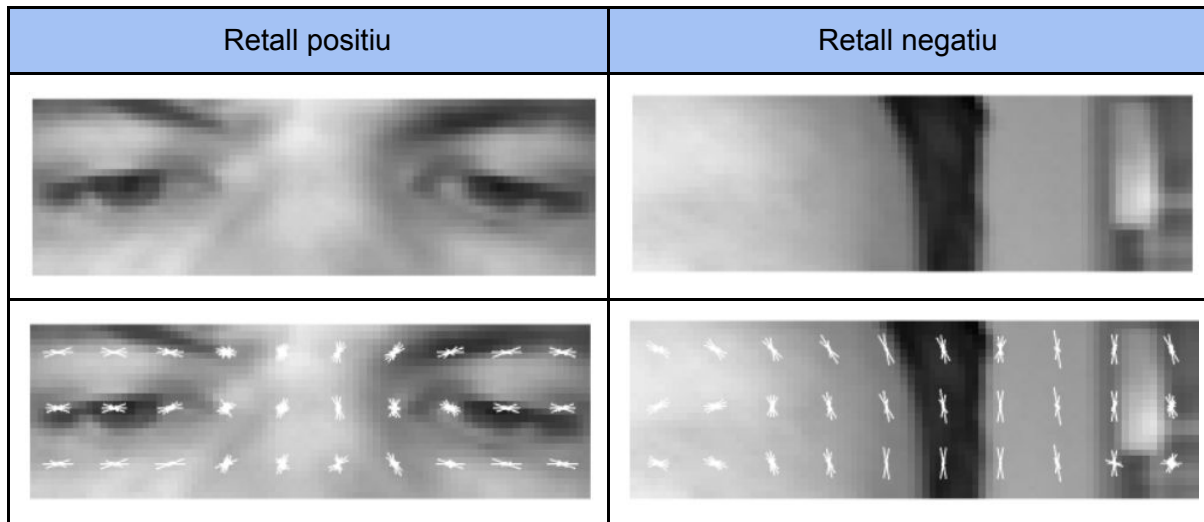


Figura 2.2. Comparativa de retall positiu i retall negatiu: característiques HOG.

A la figura anterior, on anomenem *mostra positiva* aquella que consisteix en el retall d'uns ulls, i *mostra negativa* altrament, podem comparar les característiques HOG d'ambdós retalls. Podem veure que els gradients de la imatge de l'esquerra mantenen una distribució molt característica, així que és molt difícil que una mostra negativa pugui arribar a emular aquesta distribució.

Per tant, les característiques HOG són un bon conjunt de característiques per un classificador SVM, ja que es diferencien clarament entre classes, i per això el classificador necessita un nombre tan baix de vectors de suport per establir una separació entre mostres.

Estudi estadístic

A part de les característiques HOG, hem realitzat una anàlisi d'altres característiques i llur impacte en les prestacions del classificador.

Pel que fa a descriptors de forma, vam provar l'excentricitat. Per fer-ho, binaritzem la imatge i, per cada un dels costats del rectangle, cerquem el component connectat que correspon a l'ull i n'obtenim l'excentricitat.

Tenir en compte l'excentricitat (és a dir, sumant característiques HOG i excentricitat) va reduir molt lleugerament la precisió, patint una caiguda del 0.07%. Considerem que no fou significativa.

També es van valorar les característiques SURF. Vam detectar els punts SURF i en vam extreure les característiques. Aquestes característiques van incrementar significativament el temps d'execució del programa.

L'impacte que aquestes característiques van tenir sobre la precisió fou nefast: només un 60.29% de precisió, i a més la proporció de vectors de suport fou del 79.62%. És evident que és inacceptable. Per tant, vam descartar aquestes característiques.

A continuació, vam examinar el comportament del classificador amb les característiques LBP (patrons binaris locals), tot emprant la mateixa mida de cel·la que l'anteriorment emprada amb les característiques HOG.

Aquestes característiques van demostrar millor comportament que les SURF. Concretament, amb només 193 vectors de suport (0.634%) ha aconseguit un 0% d'error. Per tant, serà indispensable valorar-ne el sobreajustament més endavant. Això ho sabrem si apliquem el classificador sobre noves imatges.

Estudi estadístic dels classificadors				
Característiques	Vectors de suport	Proporció de VS / mostres	Precisió	Matriu de confusió
HOG	139	0.457%	99.995%	28899 0 2 1519
HOG + excentricitat	186	0.611%	99.921%	28899 0 24 1497
HOG + SURF	32103	79.62%	60.29%	17336 11563 517 1004
HOG + LBP	193	0.634%	100%	28899 0 0 1521

Figura 2.3. Taula comparativa d'alguns classificadors estudiats.

Prestacions en detall					
Característiques	Precisió	Matriu de confusió	Recall 0	Recall 1	F1-score (total)
HOG	99.995%	28899 0 2 1519	100%	99.868%	99.995% 99.934% 99.964%
HOG + excentricitat	99.921%	28899 0 24 1497	100%	98.422%	99.962% 99.204% 99.583%
HOG + SURF	60.29%	17336 11563 517 1004	59.988%	66.009%	74.161% 76.871% 75.516%
HOG + LBP	100%	28899 0 0 1521	100%	100%	100.00% 100.00% 100.00%

Figura 2.4. Taula mostrant prestacions detallades d'alguns classificadors estudiats.

Per últim, vam valorar les característiques Haar, emprades a l'algorisme de Viola-Jones. Combinades amb HOG, han necessitat 1640 vectors de suport. Aquestes característiques han incrementat sobremanera el temps de computació i no han mostrat un rendiment gaire bo, amb una precisió del 65%. Al veure que cada validació suposava una penalització greu, no l'hem inclosa en aquest classificador combinada amb les altres característiques, ja que creiem que no aportaria cap benefici concret. Per exemple, al combinar-lo amb HOG en reduïa les prestacions; així, una característica que no millora el classificador no és susceptible de ser inclosa.

Seguidament, aplicarem el classificador amb una finestra lliscant sobre imatges de prova (és a dir, no incloses al conjunt de dades sobre el qual estem treballant).

Hem fet diverses proves per la finestra lliscant. Recorrent els píxels en salts de 5px, el rendiment és suficient i aconseguim un bon equilibri entre la velocitat de detecció i la precisió d'aquesta. A cada retall li apliquem el classificador i n'obtenim tant la predicció com la puntuació d'aquesta. A més puntuació, més *seguretat* té el classificador en la predicció. Conforme anem lliscant la finestra, quan trobem un retall classificat positivament, en comparem la puntuació amb la millor puntuació que hem trobat fins aquell moment. Això ho fem per discriminar possibles falsos positius del classificador. Al final de l'execució, tindrem el retall que millor puntuació ha obtingut d'entre els possibles positius.

Donada una imatge de prova, aquest és el resultat d'aplicar-ne la detecció que hem desenvolupat, amb salt de 5 en 5 píxels:

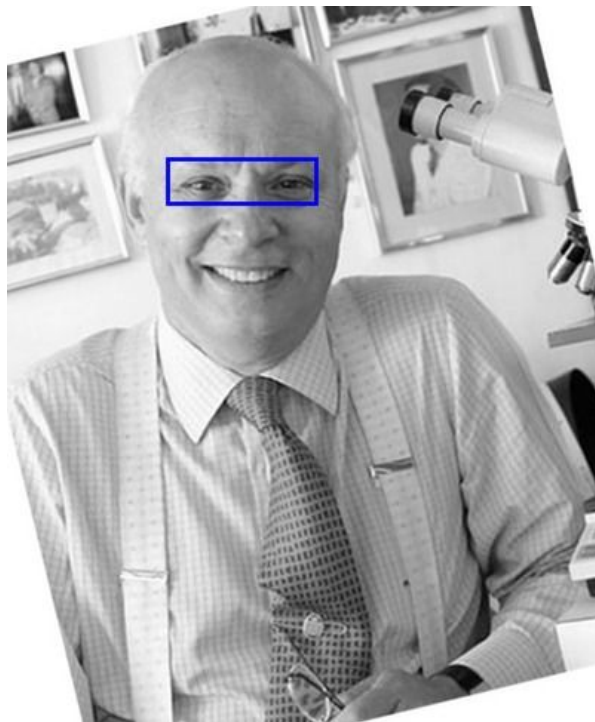


Figura 2.5. Detecció d'ulls amb finestra lliscant, amb salt de 5 píxels.

3. Detecció de mirada

Per detectar la mirada, ens serà molt útil un full de càlcul, anomenat *Miram.xlsx*, que conté les etiquetes positives (mira a càmera) i negatives (no mira) per cadascuna de les 1520 imatges de què disposem.

Podem entrenar un altre classificador SVC amb aquestes noves etiquetes.

Per validar els resultats, hem fet una *k-fold Cross Validation* amb $k = 5$. El valor d'aquest hiperparàmetre concorda amb la partició que teníem en ment originalment de 80% per les dades d'entrenament i 20% per les dades de prova.

Per cada iteració de la validació creuada, entrenem un classificador amb les dades de *train* i en predim les etiquetes per les dades de *test*. Posteriorment, comparem aquestes etiquetes amb les proporcionades pel full de càlcul, i així podem obtenir-ne l'error per iteració.

Amb tots els errors, podem calcular l'error mitjà de la validació creuada per k iteracions.

Property ▲	Value
<input checked="" type="checkbox"/> IsSupportVector	1217x1 logical
<input type="checkbox"/> Nu	[]
<input type="checkbox"/> NumIterations	2788
<input type="checkbox"/> OutlierFraction	0
<input type="checkbox"/> ShrinkagePeriod	0
<input checked="" type="checkbox"/> Solver	'SMO'
<input type="checkbox"/> Y	1217x1 double
<input type="checkbox"/> X	1217x648 double
<input type="checkbox"/> RowsUsed	[]
<input type="checkbox"/> W	1217x1 double
<input checked="" type="checkbox"/> ModelParamet...	1x1 SVMParams
<input type="checkbox"/> NumObservati...	1217
<input type="checkbox"/> Hyperparamet...	[]
<input type="checkbox"/> PredictorNames	1x648 cell
<input type="checkbox"/> CategoricalPre...	[]
<input checked="" type="checkbox"/> ResponseName	'Y'
<input type="checkbox"/> ExpandedPredi...	1x648 cell
<input type="checkbox"/> ClassNames	[0;1]
<input type="checkbox"/> Prior	[0.3969 0.6031]
<input type="checkbox"/> Cost	[0 1;1 0]
<input checked="" type="checkbox"/> ScoreTransform	'none'
<input type="checkbox"/> Alpha	537x1 double
<input type="checkbox"/> Beta	648x1 double
<input type="checkbox"/> Bias	1.6985
<input checked="" type="checkbox"/> KernelParamet...	1x1 struct
<input type="checkbox"/> Mu	[]
<input type="checkbox"/> Sigma	[]
<input type="checkbox"/> SupportVectors	537x648 double
<input type="checkbox"/> SupportVector...	537x1 double

Figura 3.1. Propietats del classificador de mirada SVC (només HOG).

A la figura anterior podem veure les propietats d'aquest classificador per una de les k iteracions.

El primer detall a tenir en compte és la probabilitat *a priori* del classificador (és a dir, calculada només a partir de les dades d'entrenament i prova): podem veure que el conjunt de dades, en aquest cas, és força més equilibrat que abans, ja que la proporció entre ulls que miren a càmera (mostra positiva) i ulls que no (mostra negativa) és del 40/60 aproximadament.

Per contra, ara disposem de moltes menys mostres que abans, ja que estem classificant el tipus de mirada dels ulls. Hom pot pensar que 1217 mostres per entrenar i 303 per classificar són insuficients per un classificador SVM.

Amb aquests dos detalls (per una banda, conjunt de dades equilibrat; per l'altra, nombre reduït de mostres), el factor determinant pel classificador serà el nombre de vectors de suport. Recordem que el límit per considerar un classificador acceptable és el 50%.

Podem veure que el SVC empra 537 vectors de suport del total (548 si n'afegim l'excentricitat com a característica a tenir en compte).

Això representa el 35.323% del conjunt de dades sencer i el 44.125% del conjunt de dades d'entrenament. Prenent la mesura més estricta, veiem que la proporció s'apropa al 50%, però encara el podem considerar acceptable.

Un cop realitzada aquesta anàlisi del nostre classificador, podem procedir a avaluar la precisió d'aquest amb coneixement de causa:

Rendiment del classificador amb 5-fold CV					
Característiques	Entrenament	Prova	Vectors de suport	Proporció de VS/mostres	Precisió
HOG	1217	303	537	44.125% (entrenament)	83.22%
HOG + excentricitat			548	45.029% (entrenament)	83.55%

Figura 3.2. Rendiment del classificador de mirada. Validació creuada, 5 iteracions.

A part de HOG i l'excentricitat, per detectar la mirada hem provat també les característiques SURF, els LBP i les característiques Haar. L'excentricitat ha demostrat millor comportament en aquest classificador anteriorment, segurament per l'especificitat de la tasca a realitzar i la relativa importància que té la forma de l'iris en la mirada. En aquest cas, el cost de computar les característiques Haar no era tan gran com a l'anterior classificador i sí que l'hem pogut incloure. No obstant, llur rendiment tampoc ha sigut molt bo.

4. Eina de proves

Per tal d'agilitzar el procés de selecció de característiques, entrenament i execució, s'ha desenvolupat una eina auxiliar. Aquesta eina permet avaluar el detector d'ulls, així com el detector de mirada, podent personalitzar les característiques a tenir en compte.

La interfície inicial és el menú següent que ens permet fer la tria entre dos opcions: entrenament manual dels SVM i validació creuada.

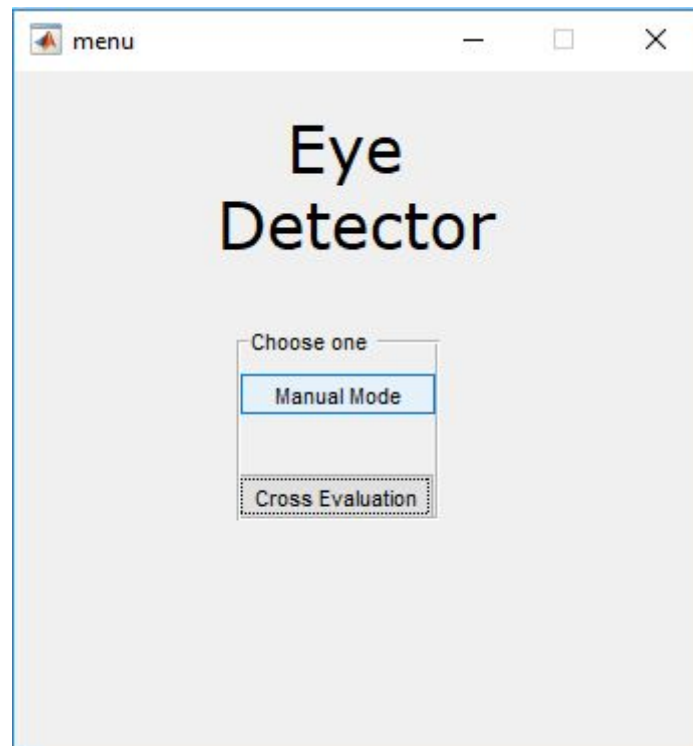


Figura 4.1. Esquema del menú principal de l'eina.

El “Manual Mode” correspon a la interfície manual de l'eina. L'aparença d'aquesta funcionalitat és la següent:

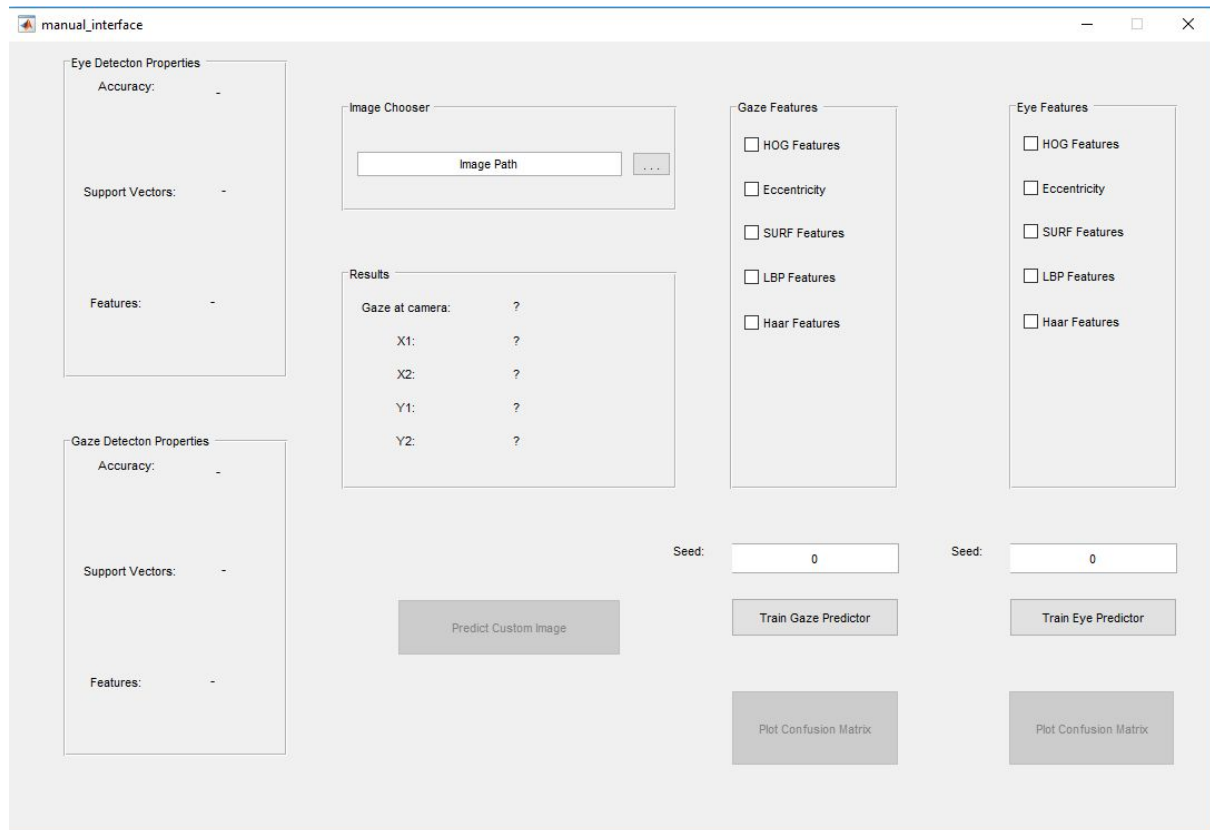


Figura 4.2. Esquema de la interfície del mode avaluació manual.

S'inclou una opció per fer servir una llavor; així, es pot replicar el comportament en futures execucions canviant les característiques emprades.

A l'esquema es poden veure les diverses seccions de l'eina:

Propietats de detecció

A l'esquerra podem veure dues seccions on visualitzar les prestacions dels classificadors d'ulls i de mirada, respectivament. Es mostra la precisió, el nombre de vectors de suport, i el nombre de característiques emprades.

Selecció d'imatge

Per avaluar el rendiment del classificador, és imprescindible poder triar una imatge externa, aliena al joc d'entrenament, per aplicar-li la detecció i classificació.

Selecció de característiques

A la part dreta de l'esquema es poden trobar sengles panells de selecció múltiple, pels dos classificadors. Es poden entrenar els dos classificadors amb subconjunts diferents de característiques: l'entrenament serà independent. Així, es pot aconseguir un nivell de perfeccionament de la detecció molt alt.

Llavor i entrenament

Per poder replicar els resultats i garantir una certa pseudoaleatorietat a l'hora d'efectuar el mostreig de les dades, és menester emprar una llavor. Per defecte és 0, però és personalitzable; això permet obtenir resultats més robustos si volem avaluar el rendiment amb diverses proves.

Posteriorment, es pot procedir a entrenar ambdós classificadors. L'eina també proporciona les matrius de confusió d'aquests.

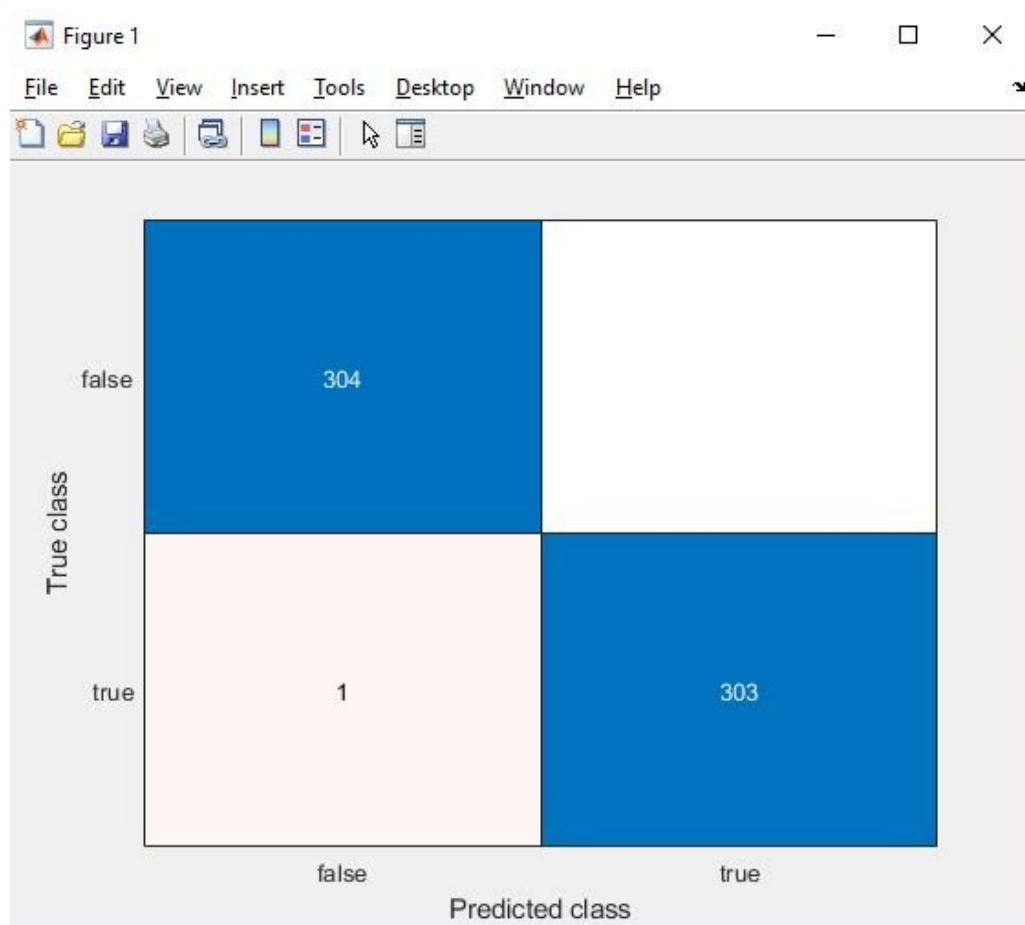


Figura 4.3. Matriu de confusió proporcionada per l'eina de proves.

Per verificar el rendiment dels classificadors, és possible aplicar directament des de l'eina una validació creuada d'aquests.

Per a poder evaluar una imatge externa al joc de proves utilitzat per l'eina, haurem de seleccionar una imatge i haver entrenat prèviament ambdós SVM (detector d'ulls i detector de mirada). Un cop realitzats aquests passos, podrem testejar-los amb la imatge seleccionada.

A més de poder seleccionar manualment les característiques a emprar, l'eina permet provar totes les combinacions possibles de característiques, accedint-hi mitjançant el menú principal a l'apartat de "Cross Evaluation"; i genera un arxiu de sortida amb les prestacions de cadascuna d'aquestes combinacions. D'aquesta manera es pot trobar el conjunt de característiques que millor rendiment permet obtenir amb els conjunts de dades.

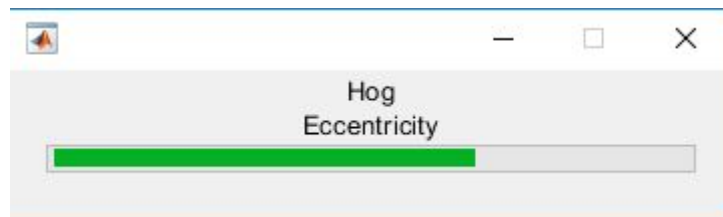


Figura 4.4. Eina provant diverses combinacions de característiques.

La interfície d'aquesta funcionalitat és la següent:

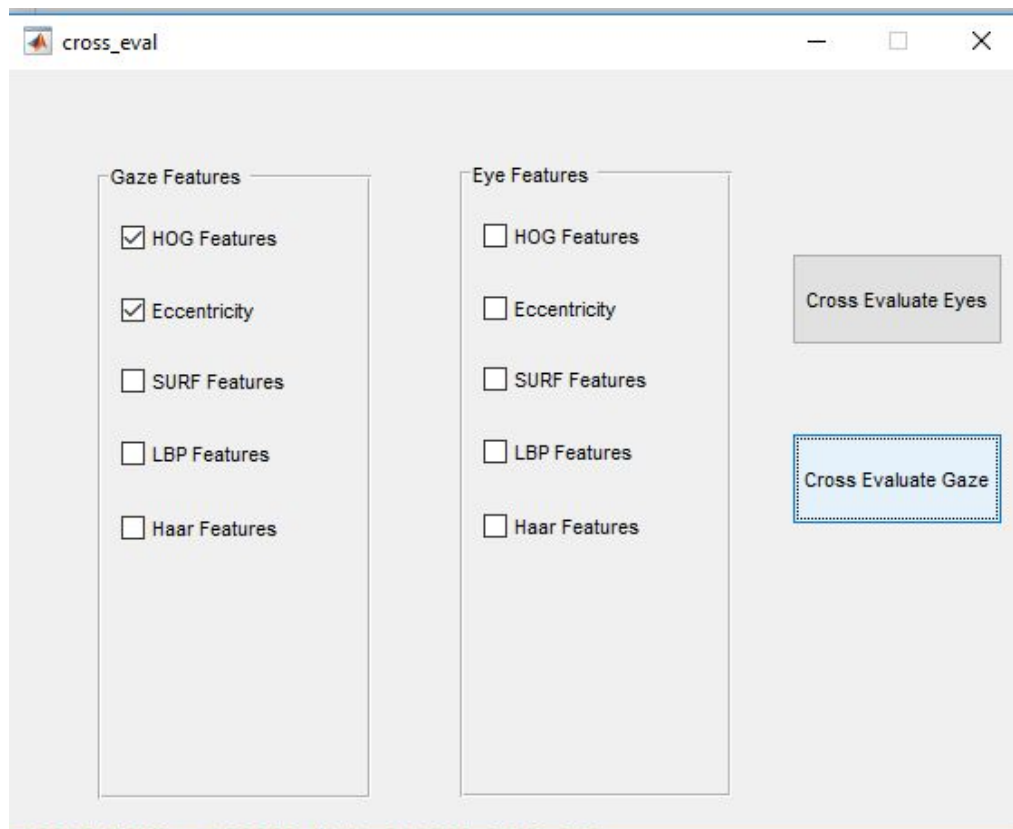


Figura 4.5. Esquema de la interfície del mode avaluació creuada

Igual que en la interfície anterior, podem seleccionar les *Features* a testejar per a cada SVM, i es comprovarà el rendiment per a totes les combinacions possibles entre els elements seleccionats, realitzant 5 avaluacions creuades per a cada combinació amb la finalitat d'aconseguir uns resultats constants i disminuir la variància en la precisió del SVM en execucions consecutives.

Resultats de l'eina de proves

Aquests són els resultats pel primer classificador. Atès que Haar donava mals resultats i trigava molt en computar-se per tantes imatges per la ingent quantitat d'imatges de mostra falses, hem prescindit d'aquestes característiques per la validació creuada amb les altres característiques.

Detector d'ulls							
HOG	Excen.	SURF	LBP	Haar	Precisió	Nombre de VS	Proporció de VS
					0.99737	126.2	0.518573%
					0.5	2434	10.00164%
					0.69704	2209	9.077087%
					0.99836	183.4	0.753616%
					0.65559	2605.6	10.70677%
					0.99803	128.8	0.529257%
					0.99868	129.4	0.531723%
					0.99901	164.4	0.675542%
					0.69013	2195.6	9.022025%
					0.9977	173.6	0.713346%
					0.99737	172.4	0.708416%
					0.9977	128.2	0.526792%
					0.99868	170.2	0.699375%
					0.99967	161.2	0.662393%
					0.9977	179	0.735536%
					0.99901	166	0.682117%

Figura 4.6. Resultats de l'eina de proves pel classificador d'ulls.

A continuació, analitzarem les millor prestacions del segon classificador, el de mirada. Aquest cop sí que hem pogut incloure les característiques Haar.

Detector de mirada							
HOG	Excen.	SURF	LBP	Haar	Precisió	Nombre de VS	Proporció de VS
					0.8506	552.0	45.39%
					0.61316	973.8	79.99%
					0.61711	983.2	80.79%
					0.82237	568.2	46.59%
					0.4875	879.0	72.19%
					0.83553	557.6	45.79%
					0.83947	547.8	44.99%
					0.84013	559.4	45.99%
					0.47763	922.2	75.79%
					0.61776	985.8	80.99%
					0.82566	569.0	46.79%
					0.52434	870.0	71.39%
					0.80921	569.0	46.79%
					0.5625	896.6	73.59%
					0.57237	1016.6	83.59%
					0.85461	555.6	45.59%
					0.84145	560.4	45.99%
					0.53947	943.2	77.59%
					0.83882	557.8	45.79%
					0.53882	934.8	76.79%
					0.51316	1013.0	83.19%
					0.81316	557.0	45.79%
					0.52303	888.2	72.99%

HOG	Excen.	SURF	LBP	Haar	Precisió	Nombre de VS	Proporció de VS
					0.58158	997.2	81.99%
					0.55329	999.6	82.19%
					0.82237	559.2	45.99%
					0.53355	937.0	76.99%
					0.55987	1003.6	82.39%
					0.52763	998.8	81.99%
					0.57237	981.4	80.59%
					0.57961	1010.4	82.99%

Figura 4.7. Resultats de l'eina de proves pel classificador de mirada.

Aquests són els resultats obtinguts al provar totes les combinacions de característiques. És important tenir en compte que els resultats surten d'una validació creuada amb 5 iteracions.

5. Conclusions

Aquest projecte ha representat un repte interessant: donada una imatge, detectar-ne els possibles ulls i posteriorment classificar-los per detectar si mira a la càmera o no.

L'objectiu a assolir ha comportat diversos obstacles a superar. La selecció de característiques és, probablement, un dels més importants. Triar un conjunt de descriptors adequat per les imatges és cabdal i, per fer-ho, és imprescindible conèixer a fons el classificador emprat; altrament, no es pot disposar d'un enfocament holístic a l'hora de valorar i comparar rendiments de diferents classificadors/detectors. Per exemple, atès que emprem un classificador SVM, la proporció entre el nombre de vectors de suport i el nombre total de mostres és un indicador a tenir en compte a més a més de la precisió obtinguda: un classificador amb millor precisió pot ser descartat a favor d'un altre si el primer disposa d'una proporció de VS major al 50% de les mostres totals. Com els nuclis dels nostres classificadors són lineals, podem ajustar l'hiperparàmetre C del SVC per penalitzar més o menys els vectors de suport i, així, obtenir un millor rendiment.

En general, podem afirmar que els histogrames de gradients orientats (HOG) i els patrons locals binaris (LBP) són les característiques que millors prestacions han proporcionat. Un detall a considerar és que els detectors i classificadors entrenats amb aquestes característiques no han mostrat un sobreajustament excessiu (hom pot pensar que amb precisions tan altes aquest era el cas), sinó que ha funcionat molt bé en diverses imatges de prova que no formen part del conjunt de dades original.

A més, podem assegurar que els rendiments no són fruit d'una execució puntual ni d'una partició intencionadament beneficiosa de les dades, ja que hem argumentat aquests rendiments amb validacions creuades, mostreigs aleatoris, i oferint l'opció de proveir el programa amb qualsevol llavor pseudoaleatòria. Aquests factors aporten una gran solidesa a tot l'estudi realitzat.

Finalment, s'ha de posar en valor l'eina auxiliar de proves que s'ha desenvolupat.

Aquesta permet aprofitar tot el potencial del projecte i facilita enormement provar diferents combinacions de paràmetres, tot podent optimitzar-ne el rendiment final.

Proporciona una interfície usable i intuïtiva, i fa possible aplicar els detectors a qualsevol imatge que l'usuari desitgi.