

DETECCIÓ DE CARES

En el projecte presentat annex a aquest document es presenten els diferents scripts per tal de realitzar les diferents tasques pròpies d'un projecte de reconeixement d'imatges. En aquest cas es presenta el reconeixement de cares de locutors de televisió, a partir de tres blocs diferenciats:

- **Bloc 1:** Permet, a través del terminal, la creació dels fitxers binaris de cada locutor amb els resultats dels vectors de la DCT. Rep com a entrada la direcció de les imatges (ja processades prèviament a partir de 'cpR2g', amb nivell de gris i detecció de cares), i la direcció de destí, amb el nom de la carpeta a crear en el *path*. Retorna, per tant, una carpeta en el lloc i amb el nom desitjat amb els diferents fitxers binaris; un per cada locutor. Permet introduir com a opció els paràmetres següents: dimensió de les imatges i número de coeficients, que per defecte reben els valors de 45x45 píxels i 78 coeficients, respectivament.
- **Bloc 2:** Realitza el reconeixement d'una imatge, que es rep com a paràmetre d'entrada, i es compara amb una base de dades d'entrenament en format binari. Aquest bloc també s'executa a través de terminal, i rep com a paràmetres d'entrada el directori de la imatge (no processada, sense detecció de cara i en escala de color) i el directori de la carpeta contenidora dels fitxers binari d'entrenament. També disposa de les opcions d'entrada de dimensió de la imatge d'entrada i número de coeficients, valors que han de coincidir amb els paràmetres dels fitxers binaris, així com l'operació realitzada sobre els vectors de la DCT. Aquesta operació pot ser la distància Euclidiana, establerta, per defecte, amb el valor '0' o la distància de Manhattan amb el valor '1'. Per tant, es retorna el nom del locutor més semblant a la imatge d'entrada.
- **Bloc 3:** Aquest últim bloc permet, a través del terminal, la creació de la matriu de confusió entre els diferents locutors. Rep com a paràmetres d'entrada el directori dels fitxers binaris d'avaluació i els d'entrenament, amb els vectors dels coeficients de de la DCT. Genera un fitxer de text en el directori de treball amb els resultats de la matriu de confusió. Aquest fitxer permet ser obert a través del programa Microsoft Excel. Com a opcions addicionals d'entrada només rep opció de l'operació a realitzar, com en el cas del bloc 2, i per tant és necessari que els vectors dels diferents directoris d'avaluació i d'entrenament hagin estat generats a partir dels mateixos paràmetres: mateixa dimensió d'imatges i mateix número de coeficients.

Els valors establerts per defecte i que es poden modificar a través de les entrades opcionals han estat testejats per tal d'oferir el millor rendiment en el reconeixement de cares. Pel que fa a les dimensions de les imatges a operar sempre obtindrem millor qualitat amb unes dimensions més grans, però també comporta una carga computacional molt més gran. Per tant s'ha optat per un valor entremig de 45x45. I pel que fa al número de coeficients de la DCT podem observar una situació similar, a diferència que amb molts més coeficients el resultat final varia molt menys, ja que la carga d'informació important es troba en els primers coeficients. Per tal d'obtenir una eficiència més alta s'utilitza per defecte el valor de 78 coeficients, que coincideix amb un valor de les diagonals de la transformada DCT.

Per altre banda també podem observar diversos scripts de funcions i classes externes, que permeten operar a cada bloc així com realitzar tasques més específiques:

- **Funcions:** Compendi de funcions relacionades amb el processat d'imatge i que són utilitzades per els altres scripts.
- **cpR2g:** Permet generar, de forma recursiva, un fitxer final amb les imatges processades (detecció de cares i nivell de gris).
- **MatConf:** Classe que permet crear, operar i generar la matriu de confusió.
- **Modelo:** Classe que permet crear, operar i generar un Model (nom del locutor amb els vectors resultants de la DCT de cada imatge).

En aquest projecte s'ha procurat realitzar gran part de les operacions de forma automàtica i a través del terminal, amb el retorn automàtic de fitxers binaris o de text.

Al calcular la matriu de confusió amb la distància Euclidiana i els valors per defecte ens ha quedat tal que:

	Agnes_Marques	Carles_Francino	JoanCarles_Peris	Nuria_Sole	Raquel_Sans	Xavi_Coral	ZNI
Agnes_Marques	54	0	1	1	0	0	3
Carles_Francino	0	69	13	0	0	3	3
JoanCarles_Peris	0	0	19	0	0	1	0
Nuria_Sole	15	2	0	135	0	0	2
Raquel_Sans	0	0	0	1	143	0	5
Xavi_Coral	0	3	0	0	0	46	1

Amb aquests valors en ha quedat una eficiència per a cadascun dels locutors:

- Agnes_Marques 0.915
- Carles_Francino 0.784
- JoanCarles_Peris 0.950
- Nuria_Sole 0.876
- Raquel_Sans 0.959
- Xavi_Coral 0.920

En canvi per a la distància Manhattan i amb els valors per defecte, queda una matriu tal que:

	Agnes_Marques	Carles_Francino	JoanCarles_Peris	Nuria_Sole	Raquel_Sans	Xavi_Coral	ZNI
Agnes_Marques	53	0	2	1	0	0	3
Carles_Francino	0	65	13	0	0	5	5
JoanCarles_Peris	0	0	19	0	0	1	0
Nuria_Sole	11	0	0	131	0	0	12
Raquel_Sans	1	0	0	2	144	0	2
Xavi_Coral	0	3	2	1	0	43	1

I les seves eficiències per a cada locutor:

- Agnes_Marques 0.898
- Carles_Francino 0.738
- JoanCarles_Peris 0.950
- Nuria_Sole 0.850
- Raquel_Sans 0.966
- Xavi_Coral 0.860

Podem observar que és més eficient utilitzar la distància Euclidiana, ja que obtenim un valor d'eficiència general de 0,900, i per a la distància de Manhattan tenim una eficiència general de 0,877.