

# VISIÓ ARTIFICIAL: CATEGORITZACIÓ AUTOMÀTICA D'IMATGES EN BASE AL CONTINGUT VISUAL

*Pilar Rosado*  
**Artista i Investigadora del grup BRAC**

*Ferran Reverter*  
**Universitat de Barcelona. Especialista en Anàlisi de Dades**

La necessitat de classificar imatges en base al seu contingut semàntic és cada vegada més important per poder indexar i recuperar informació de les grans bases de dades. Els models de visió per computador utilitzats per a la detecció de patrons en col·leccions d'imatges d'escenes naturals (paisatges, interiors, detecció d'objectes) obtenen ja resultats del 90 % d'encert en la categorització. També posen al nostre abast noves possibilitats d'aproximació a les col·leccions de les institucions artístiques, oferint un punt de vista alternatiu. Presento aquí els resultats de l'experiència realitzada amb fons de l'arxiu de Girona, analitzant les imatges en base a les seves característiques compositives, cromàtiques i organitzatives, sense necessitat de realitzar cap anotació textual sobre les mateixes. Les analogies detectades permeten establir categories sense risc de caure en interpretacions subjectives condicionades per les preferències o coneixements previs.

## **Interseccions entre mirada artística i visió artificial**

La creativitat freqüentment és acumulativa; suma, enriqueix un àmbit de treball. Sovint el creador es qüestiona les tradicions i es submergeix en les noves possibilitats que li proporcionen les tècniques, la barreja de disciplines. L'artista i la necessitat d'innovar al llarg de la història són una constant i així les revolucions tecnològiques han comportat canvis en la representació i en l'anàlisi de la realitat.

Si l'any 1990 va ser el del projecte *Genoma*, el 2013 s'han iniciat recerques multimilionàries transcendents per a l'estudi del cervell humà. Per un costat, des de els Estats Units, el projecte BRAIN (Brain Research through Advancing Innovative Neurotechnologies) pretén fer un mapa de cada neurona del cervell humà i per altre costat, des de la Unió Europea, arrenca el projecte HBP (Human Brain Project) que té com objectiu simular el cervell mitjançant supercomputadores. És segur que en els propers temps la intel·ligència artificial serà fonamental, i una font inestimable de noves eines destinades a l'extracció i producció automàtica de coneixement, de les que els artistes es podran beneficiar.

La visió per computador o visió artificial és un subcamp de la Intel·ligència Artificial que té per objectiu programar un ordinador per a que "entengui" o "interpreti" una escena o les característiques d'una imatge. En aquest àmbit concret, els investigadors s'enfronten a dos grans problemes: en primer lloc a les limitacions que suposa registrar les característiques de les imatges en un codi abstracte i en segon lloc a la dificultat d'elaborar interpretacions a partir d'aquest codi generat. Per a superar aquests inconvenients s'han creat multitud de metodologies i s'avaluen els seus rendiments.

Experts en l'àmbit de l'anàlisi de dades i de les belles arts, vam aplegar esforços a la recerca d'interseccions entre la mirada artística i la visió artificial. L'objectiu inicial va ser obtenir un sistema de catalogació i ordenació d'imatges aplicable al camp de la creació artística, que, a part de la gestió mecànica de la documentació, facilités l'establiment de nous valors i relacions entre les imatges superant formes de mirar personals i parcials.

Al proposar una recerca d'aquestes característiques, dintre de l'àmbit de les belles arts, es va plantejar la possibilitat d'extrapolar el sistema a tot tipus d'accions entorn a la creació, en les que la comparació entre imatges fóra l'objectiu principal.

La mirada polièdrica que ha aportat el fet de treballar el tema conjuntament entre especialistes de visió per computador, estadístics (concretament a les àrees d'anàlisi de dades i reconeixement de patrons) i d'artistes, ha estat clau per a la consecució dels objectius plantejats en la recerca; uns han determinat el model idoni de visió per computador, entre els molts que existeixen, i els altres han estat decisius a l'hora d'avaluar la qualitat dels resultats obtinguts. La dificultat que suposa el fet que cada disciplina tingui una terminologia pròpia s'ha superat amb el diàleg, i la col·laboració ha resultat fructífera ja que ha propiciat plantejaments que d'altra forma haurien estat difícils d'imaginar.

En l'àmbit de la categorització, assignar un descriptor textual a una imatge no sempre resulta fàcil. En resum, podríem dir que existeixen tres models principals per a anotar imatges:

1- Anotació totalment manual: és un treball pesat ja que és necessari l'esforç humà per a etiquetar cada imatge.

2- Anotació automàtica (mètodes no supervisats): el realitza el computador sense l'ajut de l'home.

3- Anotació semi-automàtica (mètodes supervisats): divideix la base de dades d'imatges a tractar en dues parts; una per a entrenament i l'altra per a validació. El conjunt d'entrenament s'etiqueta manualment de manera que la relació entre imatges i anotacions es precisa. Després, s'empren les relacions entre les imatges i les anotacions apreses pel sistema, per a generar etiquetes en el conjunt que es desitja validar.

### **La categorització de col·leccions d'imatges com a font d'inspiració**

Per realitzar els estudis que passem a descriure hem utilitzat models de visió artificial que emulen els estadis inicials de la percepció visual humana (Lowe, 2004), juntament amb models estadístic desenvolupats inicialment per a l'anàlisi automàtic de textos (mineria de textos) (Quelhas, Monay, Odobez, Gatica-Perez, Tuytelaars & Van Gool, 2005; Bosch, Zisserman & Muñoz, 2006). Aquest enfoc del tema, extrapolat a l'àmbit de les imatges, ha demostrat resultats exitosos tant en la categorització d'un conjunt d'objectes, com de diferents escenes de paisatges urbans i campestres. S'equipara el fet de descobrir el tema en un corpus, a la determinació d'una categoria d'objectes o d'escenes a l'àmbit de les imatges. El model s'aplica a les imatges mitjançant l'ús d'un anàleg visual a la paraula format pel vector de quantificació d'una regió d'imatge.

Com a exemples de les bases de dades utilitzades pels diferents autors per a la classificació d'escenes tenim la d'Oliva & Torralba (2001) amb 2688 imatges classificades en 8 categories: 360 costes, 328 boscos, 374 muntanyes, 410 camp obert, 260 carretera, 308 interior de les ciutats, 356 edificis alts i 292 carrers (Fig. 1).

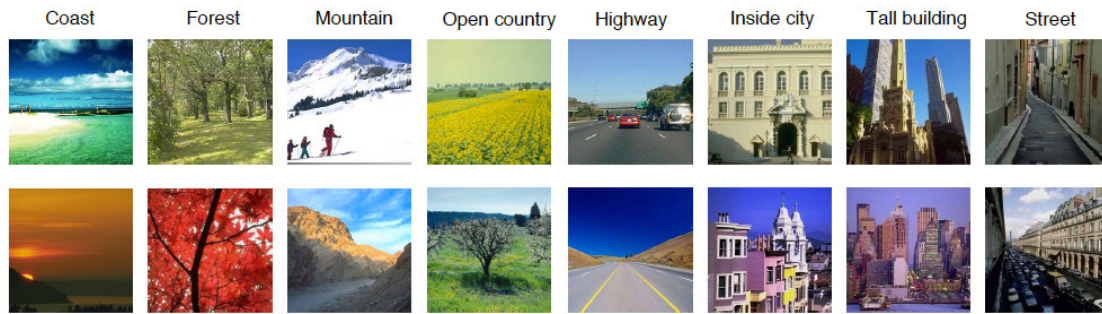


Figura 1. Exemples de la base de dades d' Oliva & Torralba (2001).

Com a exemple de bases de dades utilitzades per a la classificació d'objectes tenim la Caltech-101, amb imatges de 101 categories d'objectes i amb cada categoria d'objecte representada per una mitja de 50 imatges (Fig. 2).



Figura 2. Exemples de la base de dades Caltech-101.

En aquestes bases de dades, les categories dels objectes i de les escenes són comuns, el seu significat és compartit en general per la majoria. El repte que volíem assumir amb el nostre estudi era l'anàlisi d'imatges de contingut més abstracte, entre les que el tema, el significat o el sentit no és producte d'un acord social, sinó que es tracta de ressonàncies visuals i sincronies que l'artista creador relaciona i vincula.

La recerca es va iniciar estudiant la col·lecció de 3000 imatges digitals que l'escultor Miquel Planas utilitza en el seu procés de creació (Fig. 3). Es tracta de fotografies captades pel propi artista, la majoria d'escenes exteriors que mostren detalls des de diferents angles (fragments i particularitats que es poden considerar com elements abstractes i/o textures). Es tractava d'imatges estretes del món real, però amb intencions purament formals.



Figura 3. Alguns exemples de la col·lecció d'imatges de Miquel Planas.

Després de capturar las instantànies, l'artista utilitza la col·lecció per a la seva observació i estudi; suposen una eina de coneixement i d'accés a la seva poètica personal. Realitza sobre elles una labor taxonòmica, ordena i organitza aquestes imatges per a comprendre les relacions que ell mateix, de manera inconscient, ha establert entre elles. Al constituir una tasca personal, l'artista va percebre que amb el pas del temps, les classificacions s'anaven repetint, provocant una certa endogàmia

en els grups definits. Els resultats de les classificacions que realitzava venien en gran mesura marcats per la seva pròpia narrativa, per la seva mirada subjectiva. Aquest fet va evidenciar que les categories que establia no responien a criteris objectius i es centraven més en valors ja consolidats i reconeguts pel propi creador, molts d'ells determinats més per motius cronològics, vivencials o conceptuals de l'autor, que per la pròpia essència de les imatges captades, reduint la possibilitat de que el conjunt li aportés nou coneixement.

Paral·lelament la dificultat d'ordenar i catalogar la col·lecció amb criteris estables augmentava enormement degut a que el número d'instàncies anava augmentant exponencialment dia a dia.

En aquest sentit, una metodologia basada en aspectes computacionals, proporcionava objectivitat i era previsible que fos capaç d'establir relacions noves i, fins i tot sorprenents, entre les imatges de la col·lecció.

Vam iniciar l'estudi amb un experiment d'aprenentatge supervisat en el que un grup d'experts etiquetava manualment 5 categories d'imatges de la col·lecció que es van etiquetar com: Arquitectura Central, Pedra Geomètrica, Pedra Irregular, Pedra Texturada i Siluetes. Les 5 categories es corresponen amb 5 tipologies identificades en el total d'imatges. A la Figura 4 es mostren 5 exemples d'imatges de cada una de les categories assignades.

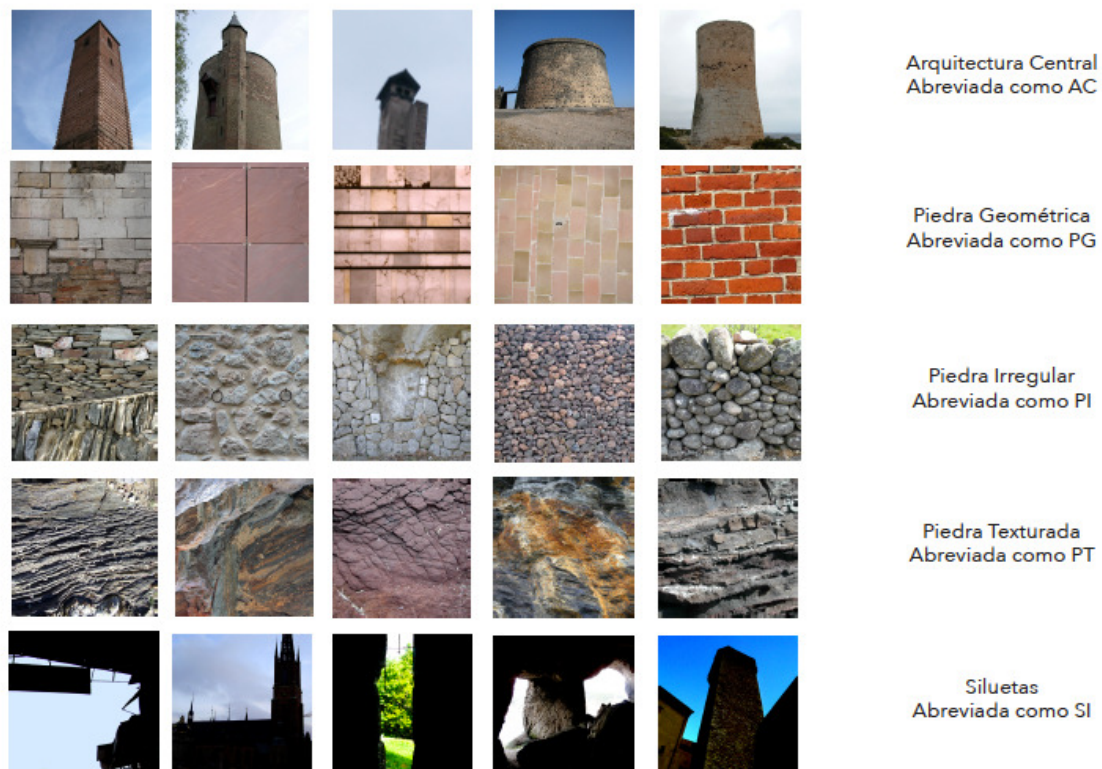


Figura 4. Cada fila mostra 5 imatges de cada una de les 5 tipologies que van definir els experts de forma manual en el conjunt d'imatges de Miquel Planas.

L'objectiu d'aquesta classificació era entrenar al sistema perquè després pogués predir la categoria a la que pertanyien altres imatges de prova. Entrenàvem al sistema amb 15 imatges de cada tipus i posteriorment el sistema categoritzava de forma automàtica la resta de la col·lecció en base a aquestes classes que havien establerts els experts (Rosado, P., Figueras, E. & Reverter, F., 2014).

Atenent als resultats que vam obtenir, es pot concloure que el model manté el bon comportament que ja havia demostrat amb imatges més comunes, quan s'enfronta a bases de dades en les que les categories existents venen determinades per continguts semàntics més abstractes involucrats en processos d'ideació i creació artística. El percentatge mitjà de classificació correcta va ser d'aproximadament el 70 % i gran part dels errors de predicció del sistema es poden justificar pel fet que la complexitat de determinació afecta directament a la fase de entrenament donat que no sempre es evident l'assignació d'una imatge a una categoria concreta, ni es simple establir els límits entre una categoria i l'altra. Analitzant els resultats que ens oferia la computadora podem observar que, moltes vegades els errors eren causats per idees preconcebudes dels experts o per problemes semàntics.

Al tractar-se d'imatges que no es corresponen amb temes usuals tals com paisatges o escenes d'interior, el coneixement previ o la experiència personal afecta notablement a l'hora de decidir si una imatge pertany a una classe o a una altra, cosa que condueix a l'establiment de unes fronteres difuses entre algunes de les classes utilitzades a la fase d'entrenament del sistema. Però també precisament per aquesta raó, el sistema es converteix en una eina d'utilitat per a l'artista, ja que li ofereix nous punts de vista diferents a la pròpia percepció.

### **Aspectes latents en el conjunt d'obra d'artistes abstractes: Antoni Tàpies**

Aquest resultat ens va conduir al plantejament de que, donada la dificultat de categorització prèvia d'algunes tipologies d'aquestes imatges, fins i tot realitzada per experts en art, cabria esperar que la classificació totalment automàtica fos més efectiva. Així va sorgir l'idea de que, més que entrenar al sistema, el veritable repte era deixar-se portar per l'algoritme amb l'objectiu que ell, sense preconcepcions, ens oferís quines categories o aspectes latents podia determinar l'estadística en una gran col·lecció.

Els artistes són generadors especialitzats d'imatges. La seva mirada recol·lecta i produeix formes i analogies de sentit a partir del seu entorn. Utilitza principis estadístics quan observen la diversitat, la processen i abstreuen el model que consideren significatiu. A les col·leccions d'imatges d'artistes hi ha uns llaços d'unió, de parentius formals que fan que constitueixin una família de significat comú.

Gràcies a l'amabilitat de la Fundació Tàpies de Barcelona, vam tenir accés a la col·lecció digitalitzada d'obra pictòrica i gràfica d'Antoni Tàpies. Utilitzant aquests models de visió artificial que emulen els estadis inicials de la percepció visual humana, hem pogut categoritzar aquesta col·lecció. Les imatges s'han analitzat en base a les seves característiques compositives, cromàtiques i organitzatives, sense necessitat de realitzar cap anotació textual sobre les mateixes.

Hem aplicat models de visió artificial sobre un conjunt de 434 imatges d'obres de l'artista de diferents èpoques i realitzades amb múltiples tècniques sobre suports variats; inclou gravats, aquarel·les, dibuixos, pintures i collages.

Alguns dels aspectes latents detectats per l'algoritme, sense cap mena d'intervenció manual, es poden consultar al "post" publicat recentment pel Museu Nacional d'Art de Catalunya (MNAC, 2016). Les analogies detectades entre les imatges de les categories visuals que hem pogut trobar a la col·lecció, ens apropen i fan més comprensibles els valors originals que el creador ha mantingut al llarg del seu període creatiu.

Amb aquest estudi suggerim la necessitat de desplegar una nova mirada sobre les col·leccions de les institucions artístiques basada en les possibilitats que posen al nostre abast les tècniques de visió artificial, i que aquest punt de vista alternatiu serveixi per incentivar, tant l'interès del públic com dels estudiosos.

El sistema que hem descrit permetria repetir els estudis sobre diferents períodes del mateix artista, o sobre col·leccions de diferents artistes o èpoques, amb els mateixos criteris. D'aquesta manera, els resultats obtinguts es podrien comparar sense risc de caure en interpretacions subjectives condicionades per les preferències o coneixements previs. I una eina de consulta d'aquest tipus, posada a l'abast del públic del museu, li permetria establir els seus propis criteris estètics de selecció.

### **Experiència amb el Fons Foto Lux de l'Arxiu Municipal de Girona**

Hem realitzat la categorització en base al contingut visual d'algunes col·leccions de l'Arxiu Municipal de Girona. En aquests exemples s'albira la visió artificial com un bon aliat per a conèixer a priori el contingut d'una col·lecció d'imatges al incorporar-la a l'arxiu, i un gran ajut a l'hora de decidir les categories o etiquetes textuais de cares a oferir-li a l'usuari que consulta un ventall més acurat de significats.

Hem aplicat aquest models sobre el Fons Foto Lux que es troba dipositat al Centre de Recerca i Difusió de la Imatge de l'Arxiu Municipal de Girona. Es tracta d'un conjunt de 96.696 fotografies generades pels fotògrafs Joan Barber i Garriga, Joan Pereferrer i Mateu i Joan Sabench en l'exercici de la seva activitat professional entre els anys 1915 i 1988. El fons Foto Lux respon en els seus orígens a la societat creada entre Joan Barber i Garriga i Joan Pereferrer i Mateu i que, amb diferents denominacions fins el 1928, s'establí a Girona pels volts de 1916 a la Gran Via i al Número 1 de la plaça de l'Oli cantonada amb el carrer ciutadans, a partir de l'estiu de 1923. La societat es dissolgué la primavera de 1934. Barber s'instal·là al núm. 29 de la Rambla de Girona i Pereferrer continuà a la plaça de l'Oli sota la mateixa denominació de Foto Lux. Pereferrer estigué en actiu fins a finals dels seixanta, quan el rellevà el seu gendre Joan Sabench, que el mantingué obert fins el 1988.

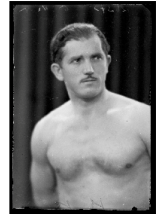
La major part del fons respon a l'activitat pròpia d'una galeria comercial, és a dir, retrats d'encàrrec de comunions, casaments, grups familiars, militars, orles, imatges sobre la ciutat de Girona i d'altres poblacions de les comarques gironines, etc.

Quan el 1989 el fons ingressà a l'Ajuntament de Girona es realitzà una primera selecció d'imatges d'una part del fons, que foren agrupades segons es tractessin de vistes urbanes, reportatges documentals o retrats i s'instal·laren en caps de arxiu convencionals. El fons està organitzat segons el contingut icònic, format i suport de les fotografies. Concretament hem treballat sobre un conjunt de 2217 imatges digitalitzades del fons de diferents tipologies.

Passem a descriure breument i mostrar algunes imatges de les categories o aspectes latents detectats per l'algoritme en la col·lecció Foto Lux. S'han analitzat les imatges en base a les seves característiques compositives, cromàtiques i organitzatives, sense necessitat de realitzar cap anotació textual. Tot i que moltes vegades resulta complicat expressar en paraules el que es percep en una imatge, intentarem descriure les qualitats que comparteixen les imatges que hem seleccionat com a representatives de cada aspecte latent.

### Aspecte 1

Trobem fotografies de retrats on el protagonista apareix en un primer terme. Són imatges en les que no apareix un marcat contrast de la figura i el fons es bastant fosc.



### Aspecte 2

Aquesta classe conté retrats de fons neutre i agrisat. Es tracta d'imatges preses des d'un pla mig en les que el rostre perd preponderància en favor del tors. La relació entre el fons i la figura s'estableix aproximadament en un 50%. La il·luminació és difusa.



### Aspecte 3

Retrats amb un fort clar-obscur, amb una font d'il·luminació molt marcada des del costat esquerre.



### Aspecte 4

Retrats a contrallum on encara s'evidencia més el contrast i el rostre passa a convertir-se pràcticament en un senzill perfil resseguit per la llum.



### Aspecte 5

Classe que agrupa fotografies realitzades en un pla de tres quarts aproximadament. En ells que pren més importància el tors. La figura està poc contrastada respecte del fons.



### Aspecte 6

Retrats de pla sencer en els que apareix tot el cos del retratat i on la persona es situa sempre davant d'una cortina amb ondulacions molt evidents. Aquest decorat es repeteix en multitud de fotografies.



### Aspecte 7

Fotografies on el marc que apareix al voltant de l'escena pren una importància considerable a nivell compositiu.





### Aspecte 8

Classe que conté imatges amb un fons principalment clar i amb detalls foscos molt contrastats que inclouen retrats però també objectes o escenes de paisatges.



### Aspecte 9

Classe que agrupa fotografies de paisatge obert amb una perspectiva elevada.



### Aspecte 10

Paisatges d'exterior amb una perspectiva normal que coincidiria amb l'alçada de la mirada del fotògraf.



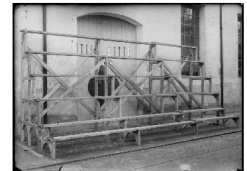
### Aspecte 11

Imatges d'interiors on es localitza un marcat punt de fuga. Les fotografies tenen una potent sensació de profunditat al contenir línies de perspectiva que convergeixen al final.



### Aspecte 12

Fotografies de paisatges urbans amb una composició eminentment horitzontal amb un punt de vista pràcticament paral·lel o molt poc inclinat.



### Aspecte 13

Imatges de paisatges on la vegetació ocupa pràcticament tota la escena. Presenten així una aparença essencialment texturada en la major part de la fotografia.



Com es pot comprovar a partir del resultats, la classificació duta a terme per la computadora sense intervenció manual dona una idea bastant rigorosa del contingut

de la gran col·lecció. Especialment al gènere del retrat s'observa com l'algoritme és capaç d'establir categories molt acurades i descriptives de la manera com han estat realitzades les fotografies: punt de vista, il·luminació, etc. Aquesta qualitat converteix al sistema en una eina de molta utilitat a l'hora de definir les categories i assignar etiquetes textuais per a cada imatge.

## **Conclusions**

En vista dels resultats obtinguts podem concloure que el sistema descrit permetria, donada una col·lecció amb gran quantitat d'imatges, realitzar una important preselecció formal agrupant-les de forma més objectiva donat que la mirada artificial no està condicionada per la percepció conceptual humana. Així, un procés d'anàlisi més objectiu com el que es presenta, aportaria nous principis d'agrupament al conjunt, que enriquiria la categorització de cares a la posterior consulta per part dels usuaris. L'accés a grans col·leccions d'imatges atenent únicament al seu contingut visual afavoriria la possibilitat d'establir categories més acurades dins de la col·lecció d'imatges.

En la actualitat disposem de tecnologies poderoses i de grans reserves de materials digitalitzats que anteriorment els estudiosos no tenien. Els mecanismes o sistemes analítics objectius, que superen formes de mirar bàsicament personals i parcials són valuosos. El model descrit estableix una categorització d'imatges, que permet una agrupació inicial basada en les sincronies formals i compositives detectades entre elles, emfatitzant els aspectes visuals i les analogies formals (compositives, cromàtiques i organitzatives) i subsidiàriament ajudaria a generar paraules clau de cadascun dels grups detectats, que repercutirien en un major enteniment de la col·lecció d'imatges afavorint també l'accessibilitat.

És impossible realitzar aquest tipus d'estudis sobre grans col·leccions d'una altra manera, per les evidents limitacions d'accessibilitat a les obres i de capacitat humana d'anàlisi. I, encara que tota categorització és per definició subjectiva en estar sustentada en un determinat criteri, la tecnologia digital ens ofereix aquesta forma d'aproximar-nos a una informació que d'una altra manera quedaria impossibilitada per la magnitud de contingut visual emmagatzemat.

## **Referències**

ARXIU MUNICIPAL DE GIRONA. <http://www.girona.cat/sgdap/cat/index.php>

BOSCH, A., ZISSERMAN, A. & MUÑOZ, X. (2007). "Image classification using random forests and ferns". In Proceedings of IEEE International Conference on Computer Vision (ICCV). Doi: 10.1109/ICCV.2007.4409066

CALTECH101. [http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/](http://www.vision.caltech.edu/Image_Datasets/Caltech101/)

FONS FOTO LUX.

[http://sgdap.girona.cat/fotoweb/Grid.fwx?search=\(\(IPTC608%20contains%20\(Fons%20Foto%20Lux\)\)%20AND%20\(FQYFN%20contains%20\(not%20fdp\)\)\)](http://sgdap.girona.cat/fotoweb/Grid.fwx?search=((IPTC608%20contains%20(Fons%20Foto%20Lux))%20AND%20(FQYFN%20contains%20(not%20fdp))))

LOWE, D. G. (2004). "Distinctive Image Features from Scale Invariant Keypoints". Int. Journal of Computer Vision, 60, 2, 91-110.

MNAC. <http://blog.museunacional.cat/la-visio-artificial-aplicada-a-les-obres-dart-una-nova-mirada/>

OLIVA, A. & TORRALBA, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175.

Doi: 10.1023/A:1011139631724

QUELHAS, P., MONAY, F., ODOBEZ, J.-M., GATICA-PEREZ, D., TUYTELAARS & VAN GOOL, L. (2005). “Modeling scenes with local descriptors and latent aspects”. *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV’05)*, 1, 883-890. Doi: 10.1109/ICCV. 2005. 152

ROSADO, P., FIGUERAS, E. & REVERTER, F. (2014). “Intersecciones entre visión artificial y mirada artística”. *BRAC-Barcelona, Research, Art, Creation*. 2(1), 1-54. Doi:10.4471/brac.2014.01

ROSADO, P., REVERTER, F., FIGUERAS, E. & PLANAS, M.A. (2014). “Semantic-Based Image Analysis with the Goal of Assisting Artistic Creation”. *Lecture Notes in Computer Science*. Vol. 8671, 526 – 533. Doi: 10.1007/978-3-319-11331-9

## **Agraïments**

Agraïm l’amabilitat de l’Arxiu Municipal de Girona que ens han facilitat l’accés a la col·lecció d’imatges del fons Foto Lux, i en especial a Joan Boades i David Iglesias per la seva disposició a facilitar la recerca i els seus comentaris constructius. També manifestem el nostre agraïment a l’arxiu de la Fundació Antoni Tàpies de Barcelona i de la Comissió Tàpies que ens han facilitat l’accés a la col·lecció d’obra gràfica, pintures i llibres d’artista d’Antoni Tàpies, i a la Dra. Eva Figueras i al Dr. Miquel Planas per la seva inestimable col·laboració.