



2.10 Les matemàtiques i les imatges

Dins el camp de la visió per ordinador, el tractament o l'anàlisi de les imatges és una de les tasques fonamentals. Moltes són les aplicacions on es té una base d'imatges, sobre les quals es necessita fer una anàlisi per a una posterior interpretació o per a poder extreure una informació que després podrem processar. En algunes aplicacions, com és el cas de les aplicacions industrials i mèdiques, s'obtenen seqüències d'imatges tridimensionals 3D. Pensem en el cas de la tomografia computaritzada (TAC), la qual ens dona informació 3D de les estructures internes del cos humà. Poder entendre aquestes imatges requereix la identificació i modelització de les superfícies dels objectes que són presents en les estructures 3D. En el cas de les imatges satèl·lit, un dels problemes està donat pel soroll present en la imatge mateixa, a causa potser de problemes de captació o transmissió. En aquest cas s'intenta fer un filtratge o preprocessament, intentant eliminar aquest soroll i conservar la informació inherent a la mateixa imatge. Un problema semblant seria el de la restauració de pel·lícules antigues que estiguin deteriorades pel pas del temps. El reconeixement automàtic de les formes és també un dels camps de la visió on l'anàlisi de les imatges és una part fonamental. Pensem en el reconeixement de la signatura o la lectura automàtica de caràcters. En el cas de la videovigilància potser es tracta d'identificar els trets principals en la fisonomia d'una persona o bé d'analitzar algun objecte que hi apareix. I no oblidem el camp de la robòtica, on l'anàlisi de les imatges en temps real que s'obtenen de l'entorn 3D en què es mou el robot, és d'una gran importància en la reconstrucció del seu entorn físic. En aquest cas, l'anàlisi o processament de les imatges seria un exemple del que anomenem *visió de baix nivell*, mentre que el problema de la robòtica aniria lligat a la *visió d'alt nivell*.

Una imatge natural o una fotografia presa del món exterior, no és més que una projecció del món tridimensional 3D sobre el pla de la imatge. Aquest món exterior és complex a causa de la mateixa estructura dels objectes i la posició que ocupen dins l'escena, a més dels canvis d'il·luminació que hi pot haver dins la mateixa escena. Com a conseqüència, les imatges de la vida real o quotidiana que es projecten sobre la nostra retina esdevenen un puzzle a causa de les oclusions i ombres a què donen lloc.

La diferència entre una fotografia o imatge natural i una imatge digital, està donada pel tipus de codificació. Per posar una fotografia dins la memòria d'un ordinador, es divideix la imatge en petits trossos quadrats, que anomenarem píxels, i dins de cada un d'aquests quadrats o píxels, li associarem un número que representa la luminància, com si fos una matriu amb nombres.

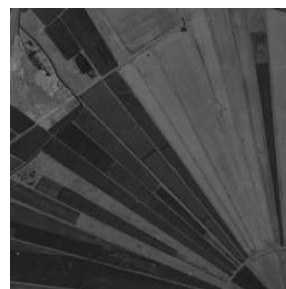
Matemàticament, una imatge o senyal, ho podem interpretar com a una funció $g(x, y)$, on (x, y) és un punt de coordenades qualsevol de la imatge i el valor $g(x, y)$ representa el nivell de gris associat al punt (x, y) . Per a la codificació, de manera semblant, cada nombre representa el nivell de gris associat al píxel (x, y) . Generalment, el negre està codificat pel zero, el 1 representa un color un poc menys negre, el 2 és encara un poc menys negre que l'1 i així. Dins la convenció que s'utilitza en el món de la informàtica, el 255 representa la codificació del color blanc.

Les matemàtiques dins el món de les imatges: modelització

Pensem per un moment en l'exemple del robot. El robot porta una càmera, la qual li dona la informació del seu entorn, però aquesta informació no és més que diferents projeccions de l'entorn 3D sobre el pla de la càmera. Dins l'anàlisi d'aquesta seqüència d'imatges, hom intenta a partir d'aquestes projeccions trobar els objectes 3D que componen l'escena, com també

conèixer la seva profunditat o distància respecte al pla de la càmera, per a saber en quin ordre el robot els anirà trobant.

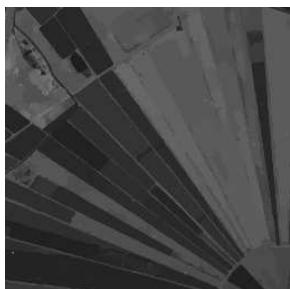
Dins un altre tipus d'aplicació, ja esmentada abans en la introducció, podem pensar en una fotografia que ens envia un satèl·lit (vegeu la imatge 1). Aquesta imatge pot tenir problemes en la mateixa captació, a causa potser del mateix moviment del satèl·lit, o bé en la transmissió (condicions atmosfèriques, etc.), incorporant una informació a la mateixa imatge, anomenada soroll, la qual és vol eliminar. Aquest soroll serà diferent (gaussiana, blanc, etc.) depenent de com s'ha produït. En aquest cas, el que s'intenta, abans de fer una anàlisi de la mateixa imatge, és un preprocessament o filtratge per a intentar separar el soroll del que es considera la informació rellevant de la imatge.



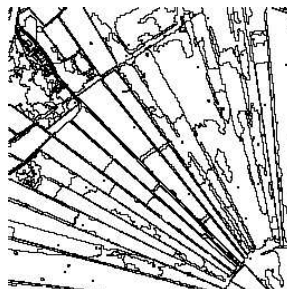
Imatge 1.

Però, quina és la informació rellevant de la imatge? Clàssicament, la detecció dels contorns, o de manera equivalent, la segmentació en regions més o menys homogènies, ha estat un dels problemes que apareixen dins la investigació de la visió per ordinador. Genèricament, les diferents regions homogènies que es poden trobar en una imatge es corresponen amb les projeccions de diferents objectes de l'escena 3D, encara que hem de tenir en compte els problemes afegits dels canvis d'il·luminació, oclusions, ombres, etc. D'aquesta forma, les regions homogènies se suposa que corresponen a parts senceres d'objectes del món real, i les fronteres als seus contorns aparents. Per regions homogènies d'una imatge, entenem regions que són uniformes respecte a alguna característica, per exemple el nivell de gris o les textures. Pensem, per exemple, en el cas de la imatge satèl·lit i posem el problema de distingir a partir d'aquesta imatge, de manera automàtica, les diferents classes de terreny que hi pugui haver: boscos, de conreu, etc. Fins i tot, ens podríem plantejar el problema de distingir entre els diferents tipus de plantacions o si dins una mateixa plantació hi ha una zona atacada per una plaga (diferent textura de les fulles).

Davant una imatge digital, un dels problemes que pretén resoldre el processament d'imatges és l'extracció automàtica de la informació de la imatge a partir dels algorismes o models existents. L'objectiu, esmentat anteriorment, seria la construcció de robots automàtics que poguessin interpretar aquesta informació, a més de millor comprendre la visió humana i animal, des d'un punt de vista biològic. Ara bé, quan volem tractar una imatge digital i per tal de definir un algorisme, com podem obtenir la informació dels contorns o fronteres de les regions homogènies? En general, els contorns que nosaltres veiem en una imatge es corresponen amb els canvis substancials en el nivell de gris (per exemple, una fotografia d'un objecte negre amb un fons blanc), és a dir, en els punts en què el valor absolut de la "derivada" sigui gran.



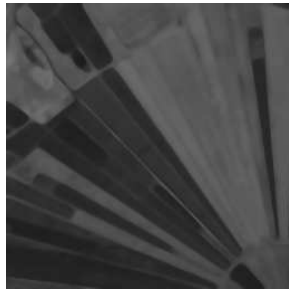
Imatge 2.



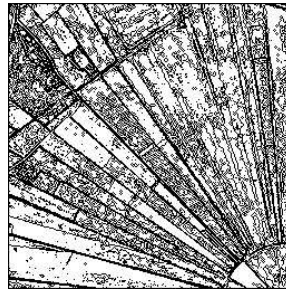
Imatge 3.

Des del punt de vista formal, la majoria dels algorismes de segmentació intenten minimitzar, per diferents camins, l'energia associada a la imatge, donada per una funció que té en compte la imatge que se segmenta i les fronteres de les regions. En l'àmbit algorímic, la idea és ajuntar dues regions veïnes si difereixen poc respecte a la característica, nivell de gris o textura, i tenir en compte que l'energia associada a aquesta nova segmentació ha disminuït. En aquest cas la segmentació inicial seria considerar que tots els píxels són diferents (cada píxel és una regió diferent), i a partir d'aquest estat començar el procés d'ajuntar les regions. Com a exemple, la imatge 2 ens dona la imatge final segmentada, amb les diferents regions homogènies, i la imatge 3 és la imatge de les fronteres d'aquestes regions, o contorns. Notem la diferència entre la imatge 1, la original, i la imatge 2, en què, per l'algorisme anterior, obtenim les regions que són semblants respecte al nivell de gris.

Tornem una altra vegada al problema de la imatge satèl·lit. En general, les zones de soroll estan associades a regions on el senyal o la imatge té freqüències altes, i en conseqüència el problema que es planteja és la seva eliminació sense destruir l'altra informació de la imatge. Dins la teoria clàssica, la imatge ha de ser filtrada (o "netejada") abans d'aplicar-li qualsevol algorisme de detecció de contorns o segmentació. En aquest cas, el que s'anomena filtratge lineal consisteix a fer una mitjana local ponderada per la funció gaussiana $G_\sigma(x) = \frac{1}{4\pi\sigma} \exp(-\frac{\|x\|^2}{4\sigma})$ o en termes més matemàtics, fer la convolució de la imatge amb la funció gaussiana. El resultat és una imatge una mica "difuminada" o filtrada, a causa de l'efecte d'aquesta mitjana ponderada. L'efecte és el mateix que el que es dona en l'aplicació següent, coneguda com l'equació de la calor: Si injectem una font de calor en un punt d'una barra de ferro, en un instant donat, llavors al cap del temps aquesta font de calor es difon al llarg de tota la barra, suposant que el coeficient lligat a la difusió és constant, fins que en un temps "gran", la temperatura en qualsevol punt de la barra serà més o menys la mateixa (efecte de difusió). L'equació de la calor és un tipus d'equació anomenada en derivades parcials, $\frac{\partial u}{\partial t}(x, t) = \frac{\partial^2 u}{\partial x^2}(x, t)$, amb la condició inicial donada per la imatge inicial $u(x, 0) = u_0(x)$ al temps $t = 0$. Si pensam en l'efecte que tendria aquest model si ho aplicam a una imatge, llavors la imatge resultant serà més o menys difosa dependent del temps que evoluciona l'equació de la calor.



Imatge 4.



Imatge 5.

L'inconvenient d'aquest filtratge donat per l'equació de la calor, o de manera semblant per la mitjana ponderada, és el fet que quan filtram la imatge, en filtrar d'igual manera els punts de contorn i els punts que es troben dins una regió més o menys homogènia, no conservam la bona localització espacial dels punts frontera. Per a tenir en compte aquest fet, existeixen unes altres equacions en derivades parcials, del mateix tipus que l'equació de la calor, les quals sobre un punt de contorn "esmorteixen" la difusió, provocant que les fronteres de les regions conservin la seva posició espacial. En canvi, a l'interior de les regions més o menys homogènies,

actuen fent la mitjana ponderada explicada anteriorment. La imatge 4 és un exemple d'aquest tipus de filtratge, en què es pot notar l'efecte de "neteja" de la imatge original.

Des d'un altre punt de vista, podríem pensar que si a cada punt de la imatge, li associem una alçada donada pel nivell de gris d'aquest punt, llavors ens podríem mirar la imatge com si fos una superfície i, per tant, la seva gràfica com un conjunt dins \mathbb{R}^3 , on $\text{graf}(u) = \{(x, y, u(x, y)); (x, y) \in \mathbb{R}^2\} \subset \mathbb{R}^3$. Si tallam la superfície de la imatge u pels plans de la forma $u = k$, $k \in \mathbb{R}$, llavors tendrem la informació sobre el relleu de la imatge: els cims, les valls, els barrancs o canvis sobtats del terreny, etc. En aquest cas podem pensar les regions homogènies com els altiplans on l'alçada en mitjana no canvia gaire, i les fronteres d'aquestes regions o contorns serien els barrancs donats pels canvis bruscs del relleu. És a dir, podem descompondre una imatge $u(x)$ amb els seus conjunts de nivell k , $\{x \mid u(x) \geq k\}$, $k \in \mathbb{R}$, o bé també amb les fronteres d'aquests conjunts de nivell. En l'exemple de la imatge 5, podem veure les línies de nivell del mapa topogràfic de la imatge satèl·lit, que són múltiples de 10. Notem que, a partir d'aquesta informació, podem reconstruir la imatge. Intuïtivament, la interpretació d'un mapa que ens mostra la topografia d'una zona més o menys muntanyosa, amb les corbes de nivell a diferents alçades, ens permet anar d'excursió, sabent a quins indrets trobarem una vall, a quins un cim, etc.

Aquesta interpretació de la imatge en termes de l'estructura del mapa topogràfic, té una bona propietat en el món de la visió: la invariància per canvi de contrast. És a dir, si u és la imatge i g és una funció no decreixent, llavors u i la composició de u amb g , $g \circ u$, tenen els mateixos conjunts de nivell. En l'àmbit de la interpretació sobre la imatge, aquesta propietat ens diu que si tenim dues fotografies d'una mateixa escena, tan sols canviant una mica les condicions de lluminositat, llavors la informació que podem obtenir de les dues fotografies és la mateixa (hi trobarem els mateixos objectes).

Altres models

Pensem, per exemple, que per un problema de transmissió d'una imatge, la imatge, o el fitxer, no ens arriba sencer. Llavors el problema que se'ns planteja és a partir de quina informació podem recuperar tota la imatge. El mateix podem pensar amb el problema del *zoom*. És a dir, a partir d'una imatge de 64×64 píxels, per exemple, com podem aconseguir una de 512×512 píxels, sense pèrdua de "qualitat". Aquests dos problemes aplicats, que fins i tot interessin en gran manera a la indústria tecnològica, estan relacionats amb el problema matemàtic de la interpolació.

Un altre tipus de problema matemàtic relacionat amb el tema de les comunicacions i visió és el de la compressió. En aquest cas els algorismes de compressió transformen les dades de la imatge en una altra representació, de tal manera que aquesta nova representació requereix menys espai d'emmagatzematge de dades. Així, el JPEG és un algorisme de compressió basat en l'anomenada transformació discreta del cosinus. Una altra eina matemàtica que ha provocat una petita revolució dins el tractament del senyal i que té una especial aplicació al camp de la representació de funcions i en particular, en el de la compressió de senyals, són les anomenades ondets.

Aquesta secció esta basada en el treball *Un model matemàtic per a l'anàlisi d'imatges* de V. Caselles, B. Coll i J.M. Morel publicat al *Butlletí de la Societat Catalana de Matemàtiques*, vol. 14, núm. 1, 1999. pàg. 63-83, i en el llibre *Digital Image Processing* de R. Gonzalez i R. Woods, Addison-Wesley, 1992.