MAT²
MATerials MATemàtics
Volum 2006, treball no. 3, 20 pp. ISSN: 1887-1097
Publicació electrònica de divulgació del Departament de Matemàtiques
de la Universitat Autònoma de Barcelona
www.mat.uab.cat/matmat

Imatges i matemàtiques

Joaquim Bruna
Jordi Saludes

1 Càmeres i imatges digitals

Darrerament les càmeres fotogràfiques digitals s'han posat de moda, ara que els preus comencen a ser més assequibles s'estan convertint en un element de consum, el regal idoni en molts casos. Les càmeres tradicionals, les que utilitzen el típic rodet, també anomenades *analògiques*, estan trobant el seu lloc en els calaixos de les coses velles (excepte per als professionals del món de la fotografia). En aquesta col·laboració parlarem d'una forma molt elemental d'alguna matemàtica que hi ha dins d'una càmera digital.

Quan ens presenten una càmera digital ens parlen d'unes característiques tècniques. Per exemple, en la càmera

Figura 1: SONY CyberShot DSC-P100

En aquesta càmera digital hi distingim les següents característiques, segons el fabricant:

- Lent *Carl Zeissl Vario-Tessar*. Zoom Òptic 3x
- Zoom digital de precisió 6x
- Resolució 5.0 megapíxels, (2592×1944)
- 14 bits per píxel
- Format d'enregistrament **jpeg**, **mpeg1**
- *Memory stick* de 8 Mbytes

El primer punt fa referència a les propietats òptiques i mecàniques de la càmera, que són les mateixes que podem trobar descrites en el cas de les càmeres tradicionals, i no en parlarem aquí, llevat d'insistir en el fet que l'òptica d'una càmera, tant si és digital com si no, és l'aspecte més essencial que en determina la seva qualitat. Considerem els conceptes dels altres punts.

Comencem amb la *resolució*. Aquest és un terme ja popular; tothom intueix que quanta més resolució, millor és la càmera (cosa que només és certa amb òptiques comparables). Que és això d'un *píxel* i la resolució 2592×1944 en píxels?. Per respondre, hem d'explicar primer com una càmera digital capta les imatges. L'element bàsic, el píxel, és un sensor molt petit que quan hi incideix la llum digitalitza la informació lumínica. Concretament, converteix cadascun dels fotons que incideix en la superficie del sensor en càrrega elèctrica, que s'acumula en un condensador mentre duri el temps d'exposició; acabat aquest, s'avalúa la càrrega acumulada en cada sensor. Pel cas d'imatges en blanc i negre, senzillament podem pensar que el píxel transforma la intensitat lumínica en una intensitat de corrent elèctric que mesura mitjançant un número. Aquest número és la intensitat en una escala de grisos, corresponent el zero al negre. Pel cas d'imatges en color la cosa és més complicada, de fet hi ha diverses teories i formes de definir i digitalitzar el color. La més habitual és el sistema RGB (roig-verd-blau), on enlloc d'un sol número en tenim tres, un per a cada canal (intensitat de R,G,B respectivament) i el color que està arribant al píxel és la "suma" d'aquests tres. Aquí ens limitarem al cas d'imatges en blanc i negre, i a partir d'ara acceptarem que la nostra càmera només fa fotos en blanc i negre.

Molt bé, doncs la nostra càmera lògicament ha de tenir molts d'aquests píxels/sensors, disposats en un quadrat o rectangle, que s'anomena el CCD (charged coupled device).

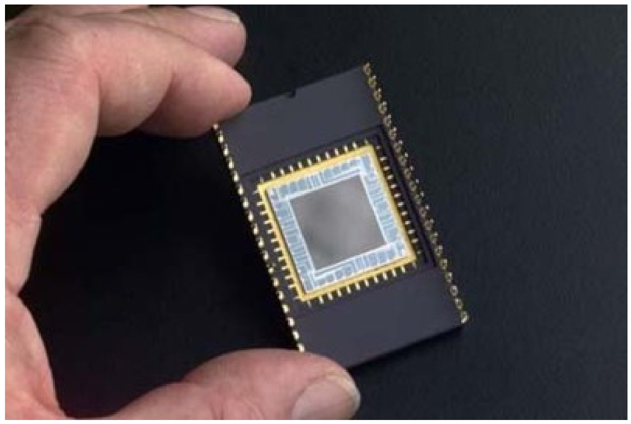


Figura 2: Un CCD

La resolució ens indica el nombre i disposició horitzontal/vertical d'aquests sensors. El 2592×1944 significa que hi tenim 1944 files, cadascuna amb 2592 píxels. Quan nosaltres, equipats amb la nostra càmera digital de resolució 2592×1944 enfoquem una escena E i fem la fotografia (en blanc i negre), que hem fet?. La part òptica del procés és la mateixa que en una càmera tradicional: en obrir el diafragma, la llum entra en la càmera a través del joc de lents i l'escena E s'hi reprodueix darrera en un cert pla. En les càmeres tradicionals, precisament en aquest pla hi ha la pel.lícula del rodet, la llum en provoca les reaccions fisico-químiques que l'impressionen i s'obté el negatiu. En les càmeres digitals, la pel.lícula es substituida pel CCD de 2592×1944 píxels, cadascun d'ells rep la informació lumínica i la transforma en un número mitjançant un conversor analògic-digital.

El següent *símil pluviomètric* ens ajudarà a comprendre millor el procés. Pensem en un camp pla R de mida 2592×1944, en metres, on hi està plovent però de forma irregular, és a dir, acceptem que en alguns punts hi plou més que en d'altres. Per mesurar això, cobrim R amb $2592 \times 1944 = 5\,038848$ (que correspon als 5.0 Megapíxels de l'anunci) dipòsits quadrats d'un metre de costat que actuen com a pluviomètres, i durant un determinat *període d'exposició*, posem una hora, recollim l'aigua que cau i mirem els litres que conté cadascun del dipòsits. Així tindrem exactament els litres que han caigut en una hora en cadascun dels $5\,038848$ metres quadrats del camp. Doncs bé, si substituim gotes d'aigua per fotons, dipòsits per píxels, l'hora per uns milisegons, tindrem una imatge prou acurada del que hem fet amb una càmera digital.

Quants més píxels, quanta més resolució hi hagi, quants més pluviòmetres hi hagi, és clar que disposarem d'una informació més rica. Els píxels es miren de fer el més petits possible per tal que la superficie del CCD sigui mínima (la probabilitat d'errors deguts a la pols etc són proporcionals a l'àrea), però no són arbitràriament petits. En aquest punt és important fer una precisió que trenca un xic la validesa del símil. Suposem que enlloc de la resolució 2592×1944 tenim una càmera amb resolució doble 5184×3888 (que no està encara en el mercat), de forma que tenim quatre cops més píxels, o quatre cops més pluviòmetres, la placa de sensors es quatre cops més grossa. Modificant la distància del pla de projecció, per mitjans òptics, també fem quatre cops més grossa la imatge. Aquest procès òptic no té un equivalent en el símil pluviomètric (no podem dilatar per quatre el camp i la pluja!); per mantenir el símil pluviomètric, pensarem que multiplicar per dos la resolució vertical i horitzontal correspon a dividir en quatre parts iguals cadascun dels pluviòmetres.

Intentem posar una mica de llenguatge matemàtic a tot això. La intensitat lumínica del feix de llum que incideix en la placa de sensors (que a través de l'òptica conforma l'escena enfocada E) seria una funció $f(x, y)$ definida en un rectangle R,$f(x, y)$ representa la intensitat lumínica en el punt ideal (x, y) (així per exemple la funció $f(x, y) = 0$ seria la imatge completament negra). Així, l'escena E es modelitza mitjançant la funció f en el rectangle R. Pensem el rectangle R dividit en $2592 \times 1944 = 5\,038848$ rectangles petits R_{ij} obtinguts dividint el costat horitzontal de R en 2592 trossets iguals i el costat vertical en 1944 trossets. Cadascun dels 5 038848 píxels correspon a un rectangle Q, i calcula un valor numèric p_Q que és una mena de mitjana en Q de la funció f:

$$p_Q = \frac{1}{\text{àrea de } Q} \int_Q f.$$

En el nostre símil pluviomètric, la integral correspon al nombre total de litres caiguts dins de Q, que dividit per l'area, dóna els litres/m^2 que han caigut dins Q en terme mig [durant el temps d'exposició], és a dir el nivell de l'aigua. Així obtenim una matriu de números de 1944 files i 2592 columnes.

Una *imatge digital en blanc i negre* a aquesta resolució seria en principi una matriu de 1944 files i 2592 columnes, per tant 5 038848 nombres corresponents als 5 038848 píxels (aproximadament els 5.0 megapíxels de l'anunci). Ara bé, no hem de pensar que cadascun d'aquests nombres varien en un rang continu, perque això fora d'una banda tecnològicament impracticable i d'altra banda exigiria una quantitat de memòria desmesurada. A la realitat aquests nombres estàn *quantitzats*. En el símil pluviomètric això correspondria a que acabem mesurant els nivells en centímetres (negligint els decimals).

A la pràctica doncs, igual que en els monitors dels ordinadors, els píxels quantifiquen els nivells de gris; això vol dir que el nombre de nivells de gris diferents ja està predeterminat dins un rang, concretament mitjançant el nombre de *bits per píxel* (bit depth), que pot ser $8, 12, 14, 24, 32, 64$ etc. A la càmera de l'anunci és 14. Qué significa això? Si atribuïm 14 bits per píxel, com que a cada bit pot haver-hi 0 o 1, el nombre total de nivells de gris diferents que podem manipular és $2^{14} = 16384$ (podem pensar en els nombres de l'1 al 16384 escrits en el sistema binari), i parlem de que el nostre monitor o càmera té 16384 colors, etc. Amb 16384 colors i una resolució de 2592×1944, el tamany de la imatge és de 8 Mbytes. Això representa que només podriem guardar una foto al *memory stick*!.

Com se solventa aquest inconvenient? Doncs establint una manera (*format de dades*) de guardar aquesta informació que s'adapti a cada imatge i tingui en compte les seves particularitats. Per entendre això, pensem en un cas extrem, una imatge a la resolució 2592×1944 completament negra; convindrem que és força ineficaç i absurd codificar 5 038848 vegades el valor zero que correspon al nivell negre, oi?. Es clar que hom pot pensar en formats que permeten guardar la mateixa informació utilitzant menys espai de memòria. Per exemple, podem mirar primer el nombre N de nivells de gris diferents que apareixen a la nostra imatge concreta, els ordenem d'1 a N i guardem en una *paleta* (o *colormap*) quins són aquests N colors, i seguidament atribuïm a cada píxel el nombre entre 1 i N que correspon al seu color. En aquesta atribució també s'utilitza l'estructura concreta de la imatge; per exemple, si les cinquanta primeres files de la matriu tenen totes el mateix color k, k entre 1 i N, (el cel d'una fotografia), resulta mes fàcil codificar aquest fet que no 50×2592 vegades el valor k.

Hi ha diversos formats d'imatges digitals que es basen en aquestes idees, de forma que sense perdre res d'informació aconsegueixen codificar la imatge en molt menys espai. El més conegut és el **bmp** (mapa de bits), i n'hi ha d'altres (**gif**, **tiff**,...). Descriure aquests darrers formats no és l'objectiu aquí, direm solament que compacten la imatge sense perdre informació (*non-loosy algorithms*). Són *formats adaptatius* en el sentit que en funció de la complexitat de la imatge s'assoleix un tamany diferent, una imatge completament negra o monocolor no requereix pràcticament res, mentre que el retrat d'un grup nombrós de persones per exemple en requerirà molt més.

És important distingir entre la *compactació sense pèrdua d'informació* d'aquests formats i un altre tipus de compactació, com la del format **jpg**, **jpg2000** o **mpeg** que si comporten una pèrdua d'informació i als que ens referirem després. Per entendre millor la diferència conceptual entre ambdós nocions de compactació utilitzem un altre cop un símil. Quan anem de vacances amb la familia amb cotxe ens trobem que no ens cap tot l'equipatge al maleter, cal compactar. El primer recurs és treure l'aire de les bosses, posar els mitjons dins les sabates, desinflar les rodes de les bicis, les pilotes; això és la compactació sense perdua d'informació, perquè no deixem res. En una segona fase, veiem que ni així, i és aleshores quan ens preguntem: però és realment necessari endur-nos això o això altre? I procedim a deixar a casa allò que ens és secundari; aquesta fora la compactació amb perdúa d'informació (la no rellevant).

En el que segueix interpretarem que la imatge digital és la matriu P de mida 2592×1944 de números que corresponen als nivells de gris, i deixem de banda el format, la manera, com aquesta informació es conserva a la memòria del ordinador.

99	135	185	249	240	252	240	255	253	253	253
68	96	135	169	246	246	254	235	254	253	253
33	68	99	147	165	235	251	255	250	252	253
10	29	90	112	136	168	242	252	246	252	252
2	14	43	90	126	155	165	209	250	252	252
7	14	27	81	120	104	182	181	203	246	254
7	13	51	93	99	134	119	149	193	200	253
15	35	55	115	63	95	117	125	179	186	204
30	54	133	109	207	103	121	122	114	175	181
47	110	115	194	243	227	103	134	130	150	159

89	189	249	245	253	254
35	99	182	247	248	253
5	44	116	183	239	252
6	26	98	135	181	238
14	69	124	109	135	187
39	119	214	187	123	147

Figura 3: Imatge i representació matricial d'un rectangle 10×11 píxels del centre de la imatge. Els valors numèrics corresponen al rang $[0, 255]$. A la dreta matriu P' corresponent a una càmara amb una resolució menor.

Si utilitzem una altra càmera de resolució diferent, per exemple 1296×972 per fer una foto de la mateixa escena obtindrem una matriu diferent P'. És convenient que pensem en aquestes matrius com a *aproximacions* de l'escena E o, equivalentment, de la funció f; aquesta escena E —la funció f definida en el rectangle R — ens és, parlant estrictament, desconeguda; idealment s'obtindria fent "arbitràriament gran" la resolució, i de fet el que en veiem mitjançant els nostres ulls seria un altre aproximació (bé persones diferents veuen detalls diferents d'una mateixa escena!). En realitat doncs, d'una mateixa escena només en coneixem aproximacions i per aquest motiu en el model matemàtic que estem introduint sovint s'anomena E —o la funció f — una *escena ideal*.

Per entendre millor aquest concepte d'imatge digital parlarem ara del *zoom òptic* i del *zoom digital*, que són coses ben diferents. El zoom òptic es una operació mecànica-òptica que serveix per enquadrar, per triar l'escena E, a la qual apliquem la digitalització explicada. Suposem que tenim una escena E i que E' és una quarta part de E, diguem la inferior dreta. Si enfoquem E' i fem la foto obtenim una matriu P' de mida 2592×1944 (totes les imatges digitals fetes per la nostra càmera tenen aquesta mida); en canvi, si fem la foto a E, obtenim P, i ens interessa només la part inferior dreta, tan sols una quarta part dels píxels hi corresponen, es tracta justament de la submatriu P'' inferior-dreta de P de mida 1296×972, que és quatre cops mes petita que P'.

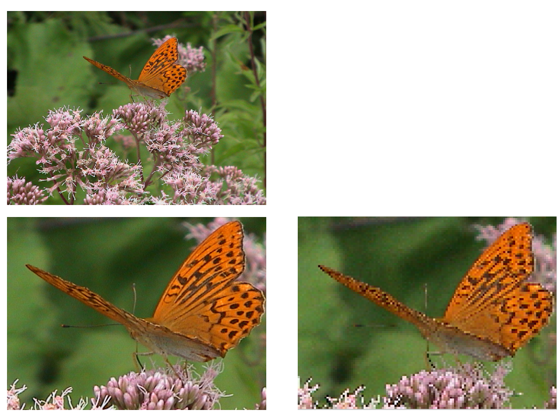


Figura 4: **A dalt**: Imatge original (P). **A baix**: Zoom òptic i zoom digital (P', P'')

Amb el zoom òptic hi tenim quatre cops mes informació, i evidentment, veurem millor els detalls de E' amb el zoom òptic. Ara bé, en molts casos no disposarem de res més que de la imatge digital P i ens interesarem pels detalls en la subescena E', que la voldríem a la mateixa mida que E (per exemple en les imatges obtingudes pels satèlits). Que fa llavors el zoom digital?. Aquesta es una operació estrictament digital, es tracta d'obtenir una aproximació de E', també de mida 2592×1944, però només a partir de

P''. Cal multiplicar per quatre la informació. Com es fa això? Pensem en el símil pluviomètric; podem pensar que cada pluviometre es pot dividir en quatre parts iguals introduint-hi dues parets perpendiculars. Si fem aquestes parets lliscants de manera que es poguin posar i treure amb facilitat, podrem comparar les "imatges" en cada resolució. Amb P'' coneixem el nivell d'aigua a cada pluviòmetre gran, i volem "endevinar" quin hagués estat el resultat si haguessim posat les parets lliscants abans de començar a ploure. Com ho fem? La manera més senzilla és dir-nos a nosaltres mateixos: "Com que hem fet tard, les posem ara" . Evidentment el resultat és que atribuim el mateix nivell a cada una de les quatre parts, perquè l'aigua es repartirà equitativament entre els nous pluviòmetres petits. Anomenem *replicació* a aquesta operació.

Ara bé, si haguessim estat a temps de posar les parets lliscants abans de ploure, és raonable pensar que aquest no hagués estat el resultat, i que una aproximació millor al nivell d'un pluviòmetre petit s'obté tenint en compte els nivells observats al seu voltant. Aquesta idea porta a altres mètodes més sofisticats; tots ells s'anomenen *mètodes d'interpolació*. En realitat, doncs, el *zoom digital inventa informació* (d'una forma adequada) a diferència del zoom òptic. Sigui quin sigui el mètode d'interpolació, el zoom digital ens obté una aproximació digital de E' de mida 2592×1944 (o d'una altra mida si es vol), que generalment no serà tan bona com ho seria P', la obtinguda pel zoom òptic. A la figura 4 podem comparar ambdues coses.

2 Processar imatges

En el model matemàtic, doncs, una imatge és una funció f en un rectangle i digitalitzar aquesta imatge significa quedar-se amb uns números

$$p_Q = \frac{1}{|Q|} \int_Q f,$$

un per a cada quadrat Q, que són la mitjana de f en Q. Aquest numeros es disposen en una matriu P. Augmentar la resolució de la imatge digital significa fer més petits els quadrats Q.

Quan a partir de la imatge digital (la matriu P guardada a la memòria de la càmera o l'ordinador) en fem una reconstrucció de la imatge (versió impresa o en pantalla LCD) el que fem és construir una funció F que és una *aproximació de f a la resolució que haguem utilitzat* del tipus

$$F(x, y) = \sum_{Q} p_Q \Phi_Q$$

on la funció Φ_Q depèn del procés. Per simplificar podem pensar en la funció característica I_Q del quadrat Q,

$$F(x, y) = \sum_{Q} p_Q I_Q$$

Dit d'una altra forma, en cada Q F és constant igual a la mitjana de f en Q. Si la resolució es pugués fer arbitràriament gran (quelcom que és imposible a la pràctica, per això f és "ideal"), és a dir, si el tamany del píxel tendeix a zero, llavors F tendeix a f.

El *processament digital d'imatges* vol dir totes aquelles operacions, transformacions, que es fan amb la matriu P de nombres de cara a una finalitat concreta. Una de les operacions més frequents és el *filtrat de soroll* , on en aquest context soroll significa totes les imperfeccions que puguin aparèixer per problemes de captació o transmissió de les imatges. La *segmentació d'imatges* té per objectiu aïllar el trets més característics d'una imatge, per exemple la fesomia d'una persona o una signatura. Es tracta en general de *detecció de contorns*. En cartografia aèria, interessa per exemple delimitar bé els llindars de les diferents finques, els camins, etc. Quan mirem per exemple una foto aèria, normalment podrem delimitar els camins i els diferents elements característics; ara bé, si això volem que ho faci un ordinador automàticament, com ho fem? També podem distingir si una certa regió ben delimitada es tracta d'un bosc, un conreu o un camp de futbol, però, novament, com s'ho fa un ordinador per esbrinar-ho? Un altre tipus de problemes està relacionat amb l'enmagatzement d'imatges. Un cas prou conegut és el del FBI als Estats Units, quan es va decidir digitalitzar tota la base de dades d'empremtes digitals, tasca que s'evidencia com inabastable sense un procés previ de compactació. *Compactar* una imatge vol dir genèricament parlant fer que la imatge ocupi menys memòria. Tot això són exemples actuals de processat d'imatges; en tots els casos, del que es disposa és d'una matriu de números d'un tamany N molt gran, en la nostra càmera $N = 5\,038848$. En termes abstractes, processar una imatge digital es fer una transformació d'aquest conjunt de mida N de números a un altre conjunt de números de mida M, és a dir, una aplicació

$$F : R^N \longrightarrow R^M$$

Quan F es injectiva, és a dir que hom pugui reconstruir la matriu $P = (p_{ij})$ a partir de la matriu $F(p_{ij})$, diem que F *no perd informació*; això significa que el procés és reversible. Per exemple *augmentar el contrast* en una imatge seria fer una operació píxel a píxel de l'estil $F((p_{ij})) = (\varphi(p_{ij}))$ on $\varphi : R \to R$ és una aplicació que augmenta el rang dels nivells de gris. Aquest és un exemple senzill; ara bé, per exemple: cóm hem de definir F, i ensenyar-ho al ordinador, per tal que la imatge processada $F(p_{ij})$ tingui els contorns resaltats? Comprimir significa que en memòria d'ordinador $F(p_{ij})$ ocupi menys que (p_{ij}), i això pot comportar pèrdua d'informació o no. Com podem comprimir de forma que l'essencial dels trets característics de la imatge es conservin i implementar-ho en un algoritme?

La resposta a aquestes preguntes *es formula en termes i llenguatge matemàtics*. Hi ha diverses tècniques en tractament d'imatges per abordar aquests tipus de problemes. En l'apartat següent presentarem d'una manera elemental una forma d'ordenar la informació d'una imatge digital, segons l'escala o resolució, que és la idea bàsica d'una teoria matemàtica molt important en el tractament d'imatges i so, *la teoria d'ondetes*, i veurem la seva utilitat en compressió.

3 L'anàlisi multiresolució de Haar

Suposem que hem fet una foto (en blanc i negre) de resolució 2592×1944. Un amic nostre té una càmera pitjor, amb resolució 1296×972, fa una foto de la mateixa escena i obté un resultat més dolent. Encara que sembli absurd, preguntem-nos el següent: *puc obtenir amb la meva càmera tot el que ell pot obtenir?* Puc obtenir la seva foto, pitjor, a partir de la meva? Dit d'una altra forma: *la meva informació conté la seva?* Posem una mica de llenguatge matemàtic i veurem no solament que si, sinó també com apareix una idea que a la vegada és senzilla i fecunda, que són les que acostumen a ser importants.

Sigui $f(x, y)$ la imatge ideal i Q un dels quadrats de la resolució baixa. Sempre podem pensar en el símil pluviometric: el meu amic ha utilitzat pluviòmetres com Q i en canvi jo ho he fet més fi, perquè he trencat en quatre parts iguals cadascun dels seus pluviòmetres, utilitzant les parets lliscants, parts que designarem Q_1, Q_2, Q_3, Q_4. El píxel corresponent de la càmera del nostre amic ha captat el valor mitjà de f en Q,

$$p = p_Q = \frac{1}{|Q|}\int_Q f$$

on $|Q|$ designa l'area de Q. En el símil, recordem que p indica el nivell de l'aigua dins Q.

Gràcies a la major resolució, en la nostra càmera aquest quadrat està trencat en les quatre parts iguals Q_1, Q_2, Q_3, Q_4 i tindrem quatre valors

$$p_i = \frac{1}{|Q_i|}\int_{Q_i} f,$$

que en el símil indiquen els nivells en els pluviòmetres petits Q_i

La seva informació consisteix dels valors p, un nombre per a cada Q, mentre que la nostra consisteix dels p_1, p_2, p_3, p_4, quatre nombres per a cada Q. Com que $|Q_i| = \frac{1}{4}|Q|$ i la integral sobre Q és la suma de les integrals sobre els Q_i (la quantitat de pluja caiguda dins Q és la suma de les quantitats caigudes dins cada Q_i), tenim que

$$p = \frac{1}{4}(p_1 + p_2 + p_3 + p_4).$$

Mirem-nos aquesta fórmula: és intuitivament evident que el nivell de l'aigua en el pluviòmetre gran és la mitjana dels nivells dels pluviòmetres petits.

Per tant efectivament a partir de la nostra informació podem recuperar la seva, això és gairebé obvi. Ara vé la pregunta clau: que és allò que *té de més la nostra informació?*, quina es *la diferència* entre ambdues? Aquesta diferència, allò que cal afegir a la imatge més dolenta del nostre amic per obtenir la nostra, la podem anomenar el *detall a la resolució* 2592×1944, és quelcom que forma part de la nostra informació però no de la seva.

Tot seguit ho formalitzem matemàticament: voldríem, per a cada Q, passar dels quatre nombres p_1, p_2, p_3, p_4 a uns altres quatre nombres, sense perdre informació (bijectiva), un dels quals sigui $p : (p, d^1, d^2, d^3)$. Els tres nombres addicionals d^1, d^2, d^3, tres per a cada Q, constituiran llavors el detall. Hi ha evidentment moltes maneres de completar p amb tres numeros de forma que aquesta assignació sigui bijectiva, però el que interessa és que no solament els tres nombres d^1, d^2, d^3 tinguin aquesta propietat, sinó també que tinguin una *interpretació* que els faci útils per a les aplicacions. Tot seguit veurem com amb uns càlculs elementals aconseguim ambdues coses.

Anomenem I la funció característica del quadrat Q, de forma que la foto que veu el nostre nostre amic a la seva pantalla LCD és

$$F = \sum_Q p_Q I_Q.$$

En canvi la nostra foto és

$$G = \sum_Q (p_1 I_{Q_1} + p_2 I_{Q_2} + p_3 I_{Q_3} + p_4 I_{Q_4}).$$

on I_{Q_i} indica la funció característica de Q_i. La diferència $D = G - F$ és el detall a la resolució 1024×1024, funció que designarem per D. Per entendre millor l'estructura d'aquest detall, pensem primer en la situació anàloga a aquesta però per a funcions d'una variable. En aquest cas Q seria un interval i en lloc de les quatre parts Q_i tindríem només les dues meitats Q_1, Q_2 de Q. Per a una funció $f(x)$ les expressions

$$F = \sum_Q p I_Q, \quad G = \sum_Q (p_1 I_{Q_1} + p_2 I_{Q_2})$$

on ara els p_Q designen mitjanes sobre els respectius intervals, representen les dues aproximacions, de forma que ara $p = \frac{1}{2}(p_1 + p_2)$. La diferència és

$$D = \sum_Q \frac{1}{2}(p_1 - p_2)(I_{Q_1} - I_{Q_2})$$

per tant vé codificada pels nombres $\frac{1}{2}(p_1 - p_2)$.

En cada Q el detall és un multiple de la funció $\Psi_Q = I_{Q_1} - I_{Q_2}$. El parell de números (dos graus de llibertat) (p_1, p_2) que porta la informació a la resolució gran en Q es descomposa en els dos parells de números $T(p_1, p_2) = (\frac{1}{2}(p_1 + p_2), \frac{1}{2}(p_1 - p_2)$. El primer és el que porta la informació a la resolució anterior, la mitjana. Les diferències de cadascun d'ells amb llur mitjana són obviament iguals però de signe oposat, i el segon número és precisament una d'aquestes diferències. Aquesta operació és reversible. La figura 5 il.lustra aquesta situació: Qualsevol parell de píxels es pot expressar com a combinació d'aquests dos parells

Reprenem ara la situació en dues variables: la informació a la resolució gran en Q la porten els quatre nombres:

p_1	p_2
p_3	p_4

Figura 5: Qualsevol parell és combinació lineal d'aquests dos.

Senzillament apliquem la transformació unidimensional anterior per files (notació $T_{\rightrightarrows}$) i per columnes (notació $T_{\downdownarrows}$)

$$
\begin{array}{lcl}
\begin{array}{|c|c|}\hline p_1 & p_2 \\ \hline p_3 & p_4 \\ \hline \end{array} & \overset{T_{\rightrightarrows}}{\Longrightarrow} & \begin{array}{|c|c|}\hline (p_1+p_2)/2 & (p_1-p_2)/2 \\ \hline (p_3+p_4)/2 & (p_3-p_4)/2 \\ \hline \end{array} \\
T_{\downdownarrows} \; \Downarrow & & \Downarrow \; T_{\downdownarrows} \\
\begin{array}{|c|c|}\hline (p_1+p_3)/2 & (p_2+p_4)/2 \\ \hline (p_1+p_3)/2 & (p_2-p_4)/2 \\ \hline \end{array} & \overset{T_{\rightrightarrows}}{\Longrightarrow} & \begin{array}{|c|c|}\hline (p_1+p_2+p_3+p_4)/4 & (p_1-p_2+p_3-p_4)/4 \\ \hline (p_1+p_2-p_3-p_4)/4 & (p_1-p_2-p_3+p_4)/4 \\ \hline \end{array}
\end{array}
$$

Figura 6: Acció de les transformacions $T_{\downdownarrows} \circ T_{\rightrightarrows} = T_{\rightrightarrows} \circ T_{\downdownarrows}$ sobre un grup de 4 píxels adjacents.

Fixem-nos en el que hem obtingut en el darrer pas: Tenim, tal com voliem, p, acompanyat de tres nombres que designarem

$$d_1 = \frac{1}{4}(p_1 - p_2 + p_3 - p_4)$$

$$d_2 = \frac{1}{4}(p_1 + p_2 - p_3 - p_4)$$

$$d_3 = \frac{1}{4}(p_1 - p_2 - p_3 + p_4)$$

i que constitueixen el detall que cercàvem. És il·lustratiu mostrar la transformació inversa que passa de (p, d_1, d_2, d_3) a (p_1, p_2, p_3, p_4) en termes gràfics

En termes de funcions, el primer quadrat té assignat la funció I_Q (és constant $= 1$ en Q), el segon la funció Ψ_1 que val 1 en la meitat superior i -1 en la inferior, el tercer la funció Ψ_2 que val 1 en la meitat dreta i -1 en l'esquerra, i finalment el quart té assignada la funció Ψ_3 producte de les anteriors. En termes d'aquestes funcions,

$$G = F + d_1\Psi_1 + d_2\Psi_2 + d_3\Psi_3,$$

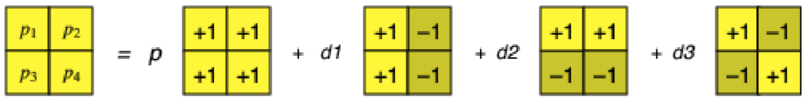


Figura 7: Qualsevol quadrat de 4 píxels es pot posar com a combinació lineal d'aquestes 4 quadrats (el vermell significa $+1$ i el blau -1). El primer element de la base dóna la mitjana. Els altres corresponen al *detall*. Observeu que el segon element de detall és el transposat del primer i que el tercer s'obté de multiplicar aquests dos primers.

i el detall és doncs $D = G - F = d_1\Psi_1 + d_2\Psi_2 + d_3\Psi_3$.

En resum, ha aparegut de forma natural la transformació lineal que associa als quatre nombres (p_1, p_2, p_3, p_4) els quatre nombres $(p_Q, d_{Q_1}, d_{Q_2}, d_{Q_3})$. Malhauradament, aquesta transformació no té una interpretació en el símil pluviomètric. Aquesta transformació de matriu A

$$A = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & -1/4 & 1/4 & -1/4 \\ 1/4 & 1/4 & -1/4 & -1/4 \\ 1/4 & -1/4 & -1/4 & 1/4 \end{pmatrix}$$

és inversible, és a dir, no perdem informació. Observem que aquesta transformació es independent del quadrat Q. Fent-la en cada quadrat Q permet descomposar l'aproximació G a la resolució 2592×1944, la foto que hem fet nosaltres, en l'aproximació F a la resolució 1296×972, la foto del nostre amic, pitjor, i el detall corresponent com a suma de tres components. Globalment estem fent una transformació bijectiva. L'aproximació F i cadascun dels detalls d_1, d_2, d_3 té la mateixa mida 1296×972.

Una altra particularitat d'aquesta matriu és que conserva la longitud dels vectors (llevat d'un factor 2): És a dir $p_1^2 + p_2^2 + p_3^2 + p_4^2 = 4(p^2 + d_1^2 + d_2^2 + d_3^2)$. Aquesta suma estesa a tots els quadrats Q s'anomena l'*energia* total de la imatge, quantitat que s'utilitza per quantificar-ne la complexitat. De fet, essencialment, la matriu A és la única matriu binària amb aquesta propietat.

Aquesta descomposició és també *independent de la resolució on comencem.* De la mateixa forma que passem de l'aproximació G a la F i als tres detalls d_1, d_2, d_3, també el nostre amic pot descomposar la seva foto F en una de pitjor i tres detalls més, i això ho podem repetir un cert nombre de vegades.

A continuació veiem un exemple d'aquesta descomposició per a una imatge concreta.

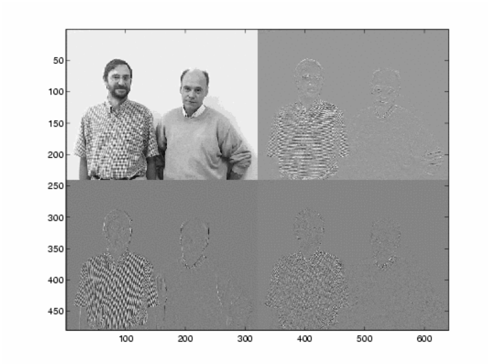


Figura 8: Exemple de descomposició de la imatge -dels autors- en les quatre components.

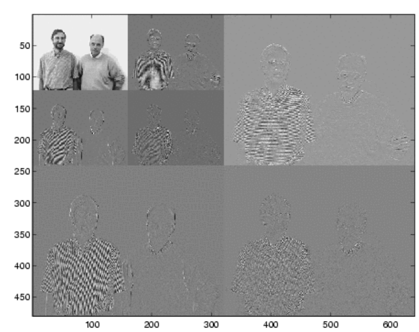


Figura 9: Idem en 2 pasos.

Aquesta descomposició tan sols depèn de la matriu A, però pot fer-se amb d'altres matrius. En el tractament digital del senyal aquest tipus d'operació s'anomena un *filtratge en quadratura*. Aquesta estructura que acabem de descriure s'anomena *l'anàlisi multiresolució de Haar*, i les funcions Ψ_i són les *ondetes de Haar*. Altres matrius A donen lloc a altres *bases d'ondetes*; la teoría d'ondetes és un dels desenvolupaments més importants en Anàlisi

Harmònica dels darrers anys, com veiem molt vinculat al tractament del senyal. Igual que per a la base de Haar, en general cada ondeta té associats dos paràmetres, un de posició que indica on està centrada, i un altre d'escala que indica la dispersió (en el cas de Haar els dos paràmetres es fonen en Q).

3.1 Per què és útil l'anàlisi multirresolució?

De qué serveix fer aquesta transformació, es a dir, organitzar la informació de partida en aquesta forma?

Comencem amb la situació unidimensional (és a dir, una fila o columna de píxels d'una imatge). La figura següent mostra com de semblant és un píxel al seu veí

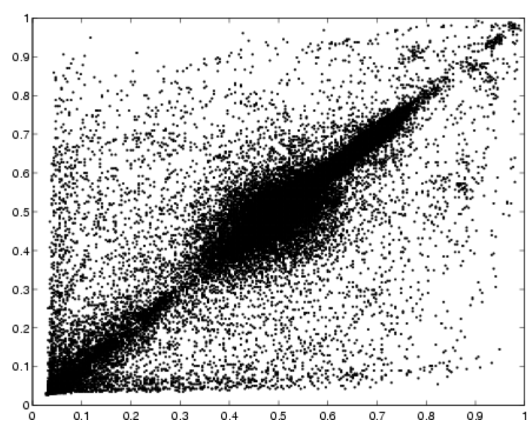


Figura 10: Gràfica del valor de p_i contra el valor del veí de la dreta (p_{i+1}). Observem que els valors es concentren a la diagonal.

Observem que el núvol de punts està concentrat essencialment a la diagonal, la qual cosa correspon a que normalment cada píxel s'assembla al seu veí. La transformació T que hem considerat abans correspon a fer un gir de 45°.

La millora que obtenim a la gràfica de la figura 11 és que els valors de la ordenada estan més concentrats (vegeu l'histograma de la figura 12).

Quelcom similar es produeix amb els detalls quan fem la transformació bidimensional.(Vegeu la figura 13)

Tenim així una transformació que, conservant l'energia total de la imatge,

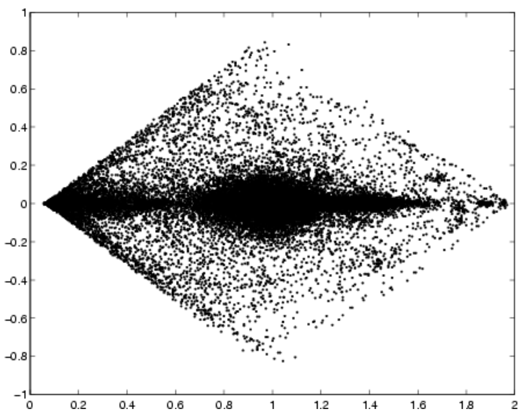


Figura 11: Gràfica girada 45° graus.

fa que tres quartes parts de la informació tingui un histograma concentrat. Per què és útil aquest fet?

En primer lloc a efectes de codificació d'aquesta informació. Que l'histograma sigui concentrat vol dir que hi ha moltes repeticions i això permet codificar aquests valors de manera eficient. Per exemple, en el cas extrem que l'histograma estés concentrat en un únic punt (un únic valor que es repeteix sempre) és evident que és molt efectiu codificar aquesta situació donant simplement el valor i el número de vegades que es repeteix.

En segon lloc, els (pocs) valors d_1, d_2, d_3 que són grans, també porten una informació rellevant. Cada coeficient per dir-ho així reflexa l'estructura de la imatge dins d'un cert quadrat Q. Doncs bé, en els llocs on no hi ha grans canvis, discontinuïtats brusques, a l'escala en qüestió, els coeficients són petits, mentre que quan hi ha una discontinuïtat forta (per exemple als contorns), un canvi sobtat, és quan el tamany dels coeficients és més gran. Si es posen a zero els valors dels detalls inferiors a un cert llindar, estem respectant les regions de la imatge on hi ha canvis, mentre que homogenitzem les zones restants, cosa que resulta útil a efectes de segmentació i compactació. L'exemple de la figura 14 mostra els resultats a partir de la foto dels autors de la figura 8.

Per aquestes dues raons, l'anàlisi multirresolució és extremadament útil a l'hora de trobar algorismes de compressió. El format **jpeg2000** es basa en aquestes idees.

Des d'un punt de vista matemàtic, la virtut principal de les ondetes en

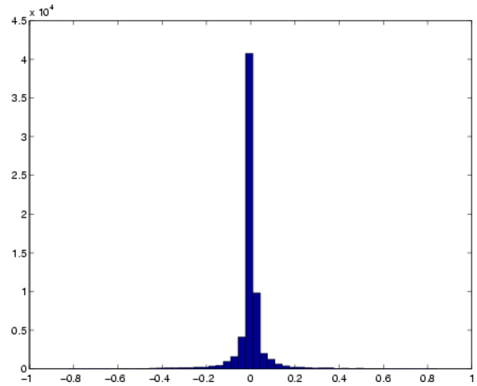


Figura 12: Histograma dels valors de l'eix d'ordenades de la gràfica anterior. Observeu que gairebé tots els punts es troben a l'interval $(-0.2, 0.2)$.

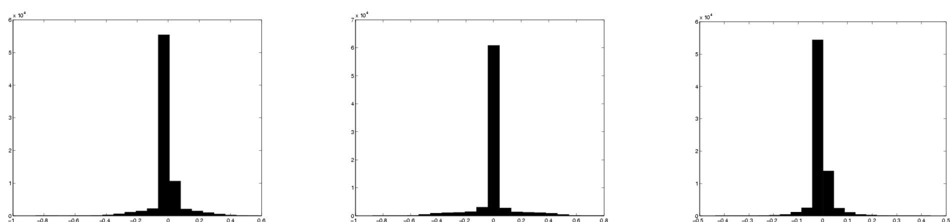


Figura 13: Histograma de d_1, d_2, d_3.

vista a aplicacions de compressió és que quan s'hi expressa una funció f(o una imatge),$f = \sum_i c_i \Psi_i$ pocs termes del desenvolupament porten ja els trets més característics. De forma més precisa, fixat un llindar (que també podria variar en cada nivell de resolució) M, si $f_M = \sum_{|c_i| \geq M} c_i \Psi_i$, per a la majoria de les imatges f, f_M és una molt bona aproximació de f.

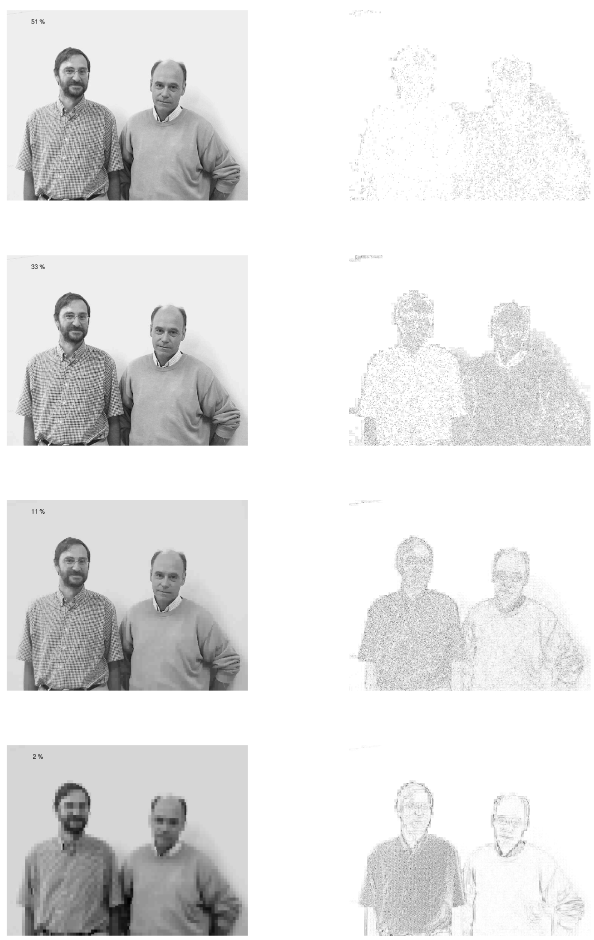
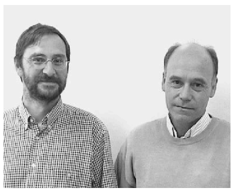


Figura 14: Resultat de comprimir posant cada vegada més coeficients a zero. **Esquerra i de dalt a baix**: Imatges comprimides amb mida que és (respecte de la imatge original): 51 % , 33 % , 11 % i 2 % . **Dreta**: Diferències entre la imatge comprimida a l'esquerra i l'original.



Joaquim Bruna
Dept. de Matemàtiques
Universitat Autònoma de Barcelona
bruna@mat.uab.cat
Jordi Saludes
Dept. de Matemàtica Aplicada II
Universitat Politècnica de Catalunya
jordi.saludes@upc.edu

Publicat el 18 de setembre de 2006