El tractament d'imatges: la beca ERC del professor Vicent Caselles

Al voltant de finals de 2012 van concedir al professor Vicent Caselles de la Universitat Pompeu Fabra (UPF) un projecte Advanced Grant titulat Inpainting tools for video post-production: Variational theory and fast algorithms. El projecte havia estat presentat en el programa Ideas de la UE dins del setè Programa marc. L'inici del projecte va ser previst per al mes d'abril de 2013 i incloïa la contractació de quatre estades postdoctorals i quatre estudiants de doctorat. La seva durada aproximada era de seixanta mesos.

Per il·lustrar els tipus de problemes que es tracten en el projecte, així com els models utilitzats, en descriurem dos: el problema d'*inpainting* d'imatges (adoptem el terme anglès) i el d'edició de vídeo a partir d'alguns fotogrames editats.

En ambdós casos l'objectiu és completar la informació en la imatge o el vídeo ja sigui perquè no la tenim o perquè la volem substituir.

Figura 1: Problema d'*inpainting*. La imatge central mostra la completesa obtinguda per una tècnica basada a completar la informació copiant segments de la part coneguda de la imatge. A la dreta es mostren les vores entre els fragments copiats.

En el problema d'*inpainting* considerem que a la imatge hi ha un forat, és a dir assumim que no tenim la imatge o dades que falten per completar-la. L'objectiu serà doncs fer-ho fent servir la informació que sí que tenim de la imatge. L'*inpainting* és una eina usualment utilitzada en programes professionals d'imatge, i les seves aplicacions més conegudes van des de treure possibles taques a la imatge o trossets de pols en un sensor d'una màquina de fotos digital, fins a treure un objecte sencer de la imatge. És important observar que l'objectiu no és retrobar el fons verdader sinó que el resultat final sembli natural.

Els mètodes que semblen produir els millors resultats s'anomenen *mètodes basats en exemplars*. La idea és assumir que les imatges tenen la propietat d'autosimilitud, és a dir, que estan formades per patrons de textura o geomètrics que més o menys es troben repetits en diferents posicions de la imatge. Amb aquesta hipòtesi es pot recuperar, per tant, la informació no disponible a partir de la informació disponible. En tot cas, el problema és determinar un mapa de còpia que especifiqui, per cada punt del forat, de quina posició de la part coneguda farem la còpia de la informació. Existeixen en la literatura models variacionals que calculen el mapa de còpia minimitzant un funcional d'energia que bàsicament imposa que la imatge final resultant del procés sigui similar a la part d'imatge disponible.

La similitud entre dues posicions de la imatge x i y es quantifica comparant la imatge en un entorn centrat de x i un atre de y. Típicament els entorns de comparació tenen tots la mateixa forma (un quadrat) i la mida es compara usant la norma L^2 (tot i haver-hi altres opcions). Així s'obtenen funcionals del tipus següent:

$$E(u,\varphi) = \int_U \int_{N(0,r)} \|u(x+h) - \hat{u}(\varphi(x)+h)\|^2 dh\, dx,$$

on U representa el forat, u és la imatge a determinar en U, $\hat{u}$ és la part coneguda de la imat-

ge, $N(0,r)$ és l'entorn quadrat de costat r centrat a l'origen del pla, i $\varphi\colon U \to U^c$ és el mapa de còpia (U^c representa la part coneguda de la imatge). Vegeu per exemple [3, 2, 1].

Per tal de conèixer millor l'acció dels mètodes descrits és interessant analitzar la regularitat dels mapes de còpia que minimitzen energies com aquesta. Assumint que la part coneguda de la imatge $\hat{u}$ és acotada i de variació acotada, és possible provar que existeixen mínims de E en els quals el mapa de còpia és límit uniforme de funcions de variació acotada que prenen infinits valors. No obstant això, les solucions calculades numèricament per imatges discretes, que en molts casos corresponen a translacions a trossos, admeten una regularitat molt més forta (com és el cas de la figura 1).

Deixem per un moment el problema de l'*inpainting* d'imatges per parlar d'edició de vídeos (vegeu figura 2). Un usuari ens dóna un vídeo en què la superfície visible d'un objecte ha estat editada en un o dos fotogrames (suposarem que és el primer o últim). L'objectiu és propagar aquesta nova informació a través del vídeo, mirant que el resultat sigui temporalment i espacialment coherent i consistent.

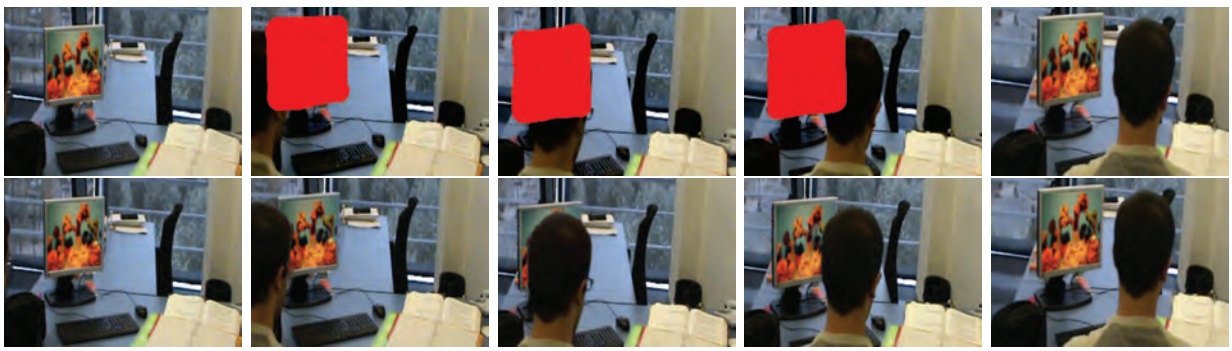

Figura 2: Alteració d'un objecte en un vídeo. En aquest exemple l'usuari modifica la imatge sobre el monitor de l'ordenador en el primer i últim fotograma i marca una regió d'edició. L'objectiu de l'algorisme és doncs propagar la modificació al dels fotogrames restants en la regió d'edició.

Consistència temporal vol dir que l'edició ha de respectar el moviment dels objectes de l'escena. *Consistència espacial* vol dir que no hi ha discontinuïtats en la frontera espacial del domini editat en cada un dels fotogrames. Observem que en el cas d'edició d'imatges la consistència espacial s'obté de manipular els gradients de la imatge (resolent l'equació de Poisson, [4]) en lloc de manipular directament els colors. Així, la mateixa idea pot aplicar-se al vídeo propagant al llarg de les trajectòries del moviment la informació codificada en els gradients dels fotogrames editats [5].

Com hem vist, els mètodes descrits en els problemes d'*inpainting* d'imatge i edició de vídeo comparteixen la idea d'explotar la redundància d'un senyal (o d'una imatge en el cas del vídeo) per interpolar les dades que ens falten en una regió. La diferència rau en el fet que la font de redundància considerada en cada cas és de diferent naturalesa.

En el cas del vídeo, la principal font de redundància és la continuïtat temporal dels objectes que en formen part, cosa que ens permet una formulació local; en altres paraules, el color d'un píxel en instants successius de la seva trajectòria varia poc. La condició local del problema ens permet modelar-lo de forma natural mitjançant operadors diferencials locals que donen lloc a EDP de propagació o interpolació al llarg de trajectòries.

En contraposició, en imatges estàtiques, la principal font d'autosimilitud ve donada per la repetició de patrons en diferents posicions de la imatge. Conseqüentment aquestes posicions poden ser diferents entre si i per tant les tècniques locals no funcionen.

Els objectius principals del projecte són (1) la investigació matemàtica dels principis de suavitat i autosimilitud que juguen en la generació d'imatges naturals, (2) la formulació matemàtica i la unificació d'ambdues idees en una única variacional, i finalment (3), aplicar-ho al

desenvolupament de models i algorismes per processar millor les imatges.

La investigació proposada se centra en la formulació i l'anàlisi matemàtica de principis variacionals per al processament d'imatges i vídeo, i en l'anàlisi de la teoria de la mesura geomètrica i equacions en derivades parcials subjacents en aquests models. En concret, un dels principals objetius és unificar la visió local *versus* la visió no local del problema, així com les corresponents expressions matemàtiques de les idees de regularitat i autosimilitud. El treball d'investigació estarà guiat per la investigació del problema d'*inpainting* (en imatges, vídeos, vídeo monocular i vídeo estèreo).

Malauradament, a causa de la trista pèrdua del professor Vicent Caselles, recentment traspassat al mes d'agost, l'execució efectiva del projecte ha quedat truncada.

Referències

[1] P. Arias, V. Caselles i G. Facciolo. «Analysis of a variational framework for exemplar-based image inpainting». *Multiscale Modeling & Simulation*, 10(2):473–514, 2012.

[2] J.-F. Aujol, S. Ladjal i S. Masnou. «Exemplar-based inpainting from a variational point of view». *SIAM J. Math. Anal.*, 42(3):1246–85, 2010.

[3] L. Demanet, B. Song i T. Chan. «Image inpainting by correspondence maps: a deterministic approach». *Applied and Computational Mathematics*, 1100:217–50, 2003.

[4] P. Pérez, M. Gangnet i A. Blake. «Poisson image editing». *ACM Transactions on Graphics*, 22:313–318, 2003.

[5] R. Sadek, G. Facciolo, P. Arias i V. Caselles. «A variational model for gradientbased video editing». *International Journal of Computer Vision*, 103(1):127–162, 2013.

Grup de Processant d'Imatges
Universitat Pompeu Fabra

Big data: la revolució de les dades

Cada dia generem una quantitat immensa de dades sobre nosaltres: amb qui parlem per telèfon, quant de temps mirem la televisió, a quines hores comprem, quins són els llibres que més ens agraden, quins medicaments prenem, quines pàgines web visitem, etc. Totes aquestes dades, que anys enrere no es podien guardar ni, pràcticament, tractar, avui dia es poden guardar i analitzar amb facilitat, i permeten l'obtenció d'informació molt valuosa. De les grans quantitats de dades i les tècniques per emmagatzemar-les i gestionar-les se'n diu *big data*, i de la disciplina que s'encarrega de la seva modelització i anàlisi, *machine learning*.

El naixement del món *big data* ha estat possible gràcies a dos factors principals. D'una banda, les millores tecnològiques, que han incrementat de manera dràstica la capacitat de generar i guardar grans quantitat d'informació a un cost molt més baix. Aquest fet ha anat acompanyat de la descentralització de la computació vers el núvol i de la substitució dels servidors centrals, cada cop més potents i cars, per xarxes d'ordinadors utilitaris, més barats i flexibles. D'altra banda, l'augment de capacitats tecnològiques ha propiciat l'aparició de nous paradigmes de tractament de dades: nous sistemes de fitxers i bases de dades, com les bases de dades NoSQL; i nous models de programació en paral·lel, com la tècnica *map reduce*. Les bases de dades NoSQL es basen en un sistema molt flexible d'emmagatzematge i permeten guardar dades no estructurades o tabulades. Això fa possible guardar dades heterogènies ràpidament i facilita la reutilització de la informació per casos que no s'havien previst inicialment. El model *map reduce* ofereix un *framework* senzill pensat per executar programes altament paral·lelitzables sense que l'usuari hagi de ser un programador