

Les mathématiques de l'image numérique

Lionel Moisan

Professeur de Mathématiques

Université René Descartes - Paris 5

L'image numérique fait aujourd'hui partie de notre quotidien, et sa production n'a jamais été aussi facile. Webcams, appareils photo et caméscopes numériques, scanners, tous ces dispositifs électroniques désormais usuels permettent à un particulier de produire massivement de telles images. Au-delà de ces utilisations domestiques, la médecine, l'astronomie, la météorologie, etc. sont aussi de grands consommateurs d'images numériques.

Le passage de la technologie analogique (film négatif recouvert d'un sel d'argent) à la technologie digitale, possède, entre autres avantages, celui de transporter la chaîne de traitement de l'image (impression, transmission, affichage, recadrage, retouche, ...) du monde optico-chimique vers celui de l'informatique, beaucoup plus souple, puissant et facile d'accès. Mais derrière cette apparente facilité se cachent de puissants outils mathématiques, opérant (heureusement) à l'insu de l'utilisateur.

L'image numérique

La photo numérique ne diffère *a priori* que très peu de son homologue argentique : le négatif photo est simplement remplacé par un capteur constitué d'une grille de cellules qui mesurent la quantité de lumière qui les frappent. L'électronique de l'appareil produit ainsi une image numérique, c'est-à-dire un tableau dont chaque cellule (appelée pixel pour *picture element*) se voit attribuer un nombre codant l'intensité lumineuse (voir encadré). Ce type de dispositif produit des images en niveaux de gris (communément appelées "images noir et blanc"), mais il peut être facilement adapté pour produire des images en couleur par l'intrication dans la grille de cellules sensibles aux trois couleurs fondamentales (rouge, vert, bleu), dont la combinaison permet de reproduire toute la palette des couleurs.

Numérisation d'une image et théorie de Shannon

L'étape de numérisation que nous venons de décrire doit respecter une règle donnée par la théorie de Shannon : la distance entre deux cellules adjacentes du capteur doit être de l'ordre de la moitié de la taille du plus fin détail que peut transmettre le système optique. Si cette règle n'est pas respectée, l'image numérique ne

reproduit pas fidèlement la réalité, et présente un aspect général flou si la distance inter-cellules est inutilement petite, ou au contraire *aliasé* si la distance est trop grande (voir encadré). Le cadre mathématique naturel pour analyser ces défauts et énoncer rigoureusement la théorie de Shannon est celui de l'analyse de Fourier, qui permet de décrire une image en termes fréquentiels.

Transformations géométriques : zoom, rotation, réduction

La théorie de Shannon va au-delà du processus d'acquisition : elle décrit comment reconstruire exactement l'image physique (continue) formée initialement sur le plan du capteur à partir de l'image numérique (discrète) enregistrée. Dès lors, il est possible de réaliser des transformations géométriques sur une image numérique de façon très précise, en simulant l'acquisition d'une nouvelle image selon une autre grille de cellules, de pas plus fin (cas du zoom, qui augmente le nombre de pixels), ou dont les axes sont tournés (cas de la rotation). Ces transformations peuvent être réalisées en pratique à l'aide d'un algorithme astucieux appelé transformée de Fourier rapide. Une autre possibilité, plus souple, est donnée par la théorie des fonctions splines, qui approchent de façon arbitrairement précise la reconstruction donnée par la théorie de Shannon, pour un coût algorithmique moindre (voir encadré).

Le cas de la réduction d'une image (diminution du nombre de pixels pour un cadrage identique) est particulièrement illustratif. En effet, la théorie de Shannon précise exactement le type de détails qui, faute de pouvoir être correctement reproduits dans l'image réduite, doivent être supprimés lors du processus de réduction, par un procédé appelé *filtrage passe-bas*. Par exemple, réduire une image d'un facteur 4 en ne gardant qu'un point sur 4 selon chaque axe (le nombre de pixels initial est donc divisé par 16) est une technique particulièrement maladroite, malheureusement encore employée par beaucoup de logiciels courants de manipulation d'images. En moyennant les 16 intensités mesurées dans chaque carré de 4×4 pixels, on obtient une réduction de qualité sensiblement meilleure (voir encadré), bien qu'encore non optimale.

Augmenter la netteté

Améliorer la qualité d'une image fait partie des opérations courantes exigibles d'une chaîne de traitement numérique. L'amélioration la plus spectaculaire est sans aucun doute le *déflouage*, qui vise à augmenter la netteté d'une image. En général, ce procédé est utilisé pour compenser l'effet de flou produit par le dispositif d'acquisition, et parfois encore accentué par de mauvaises conditions de prise de vue (autofocus défaillant, "bougé" de l'opérateur, ...). Dénommé dans les logiciels du commerce par des termes comme "renforcement", "accentuation", ou "contours plus nets", il fait partie d'une vaste classe de problèmes mathématiques appelés *problèmes inverses*. Si l'on note X l'image non dégradée (inconnue), et F l'image floue (effectivement observée), alors le passage de X à F se décrit assez simplement : l'intensité en chaque pixel p de F s'obtient comme une moyenne pondérée des intensités de X observées autour de p . Mathématiquement, cette opération est une *convolution*, et le problème du déflouage consiste à retrouver X à partir de l'observation du résultat de la convolution, F , d'où cette appellation de "problème inverse" (ici, une déconvolution).

Problèmes inverses et régularisation

Il n'est pas possible de résoudre directement ce type de problème inverse, car même lorsque la solution X associée à F est exactement calculable et unique, le procédé est instable : une petite erreur de mesure sur F (inévitables en pratique) provoque en général une perturbation énorme de X . Dans le cas des images, on constate immédiatement ce phénomène sous la forme d'une amplification du bruit de mesure lorsque l'on essaie de déflouer directement une image (voir encadré; sous Photoshop, appliquer le filtre *renforcement*—*accentuation* avec un gain élevé et un seuil nul).

Un remède à l'instabilité des problèmes inverses, appelé *régularisation*, fut proposé par Tikhonov en 1963. Il consiste à chercher une solution seulement approchée mais très régulière (ou, de façon équivalente, très plausible, si l'on réinterprète le procédé de régularisation en termes probabilistes). Mathématiquement, cela revient à minimiser une certaine énergie, et fait appel aux techniques du calcul variationnel et des équations aux dérivées partielles.

Dans le cas du déflouage, on doit pénaliser les images X oscillantes, et la pénalisation la plus pertinente connue à l'heure actuelle est la *variation totale* de X , introduite par Rudin et Osher en 1992. Elle peut être estimée numériquement en additionnant le contraste (valeur absolue de la différence des intensités) entre tous les couples de pixels adjacents de l'image. Ce procédé est très utilisé pour la restauration des images numériques (déflouage et débruitage), et fait en-

core l'objet d'intenses recherches.

Un vieux rêve : la vision artificielle

Les problèmes que nous venons d'évoquer concernent le *traitement* d'images, c'est-à-dire la transformation d'images destinées à être visualisées par un opérateur humain. Dans les années 1970, la révolution informatique promettait l'avènement rapide de l'*intelligence artificielle*, et en particulier de la *vision artificielle*, devant mener à la création de robots intelligents capables d'analyser leur environnement visuel et de se mouvoir de façon autonome. Même si cet objectif n'est pas encore atteint, le domaine de l'*analyse d'images* s'est montré mathématiquement fécond.

Détection de contours

Les premiers travaux en analyse d'images révélèrent très vite à la fois l'importance perceptuelle des contours, qui bordent les différents objets présents dans l'image, et la difficulté à modéliser leur détection. Certes, ces contours sont essentiellement constitués de pixels au voisinage desquels l'intensité présente de fortes variations (à cause du contraste brutal entre l'intensité de l'objet et celle du fond sur lequel il se découpe), mais de telles zones contrastées sont omniprésentes en pratique, notamment parce qu'une image présente des structures (et donc des contours) de toutes tailles. Comme le remarquait David Marr, pionnier américain de la vision artificielle, "*The common and almost despairing feeling of the early investigators [...] was that practically anything could happen in an image and furthermore that practically everything did.*"¹

Sur une image numérique, on peut estimer le contraste en chaque pixel par la différence entre le maximum et le minimum des intensités rencontrées sur un voisinage de ce pixel (par exemple le pixel lui-même et ses 4 pixels adjacents). Cette estimation n'est autre qu'une discrétisation de la norme du gradient de l'image continue. En repérant les pixels de l'image pour lesquels ce contraste prend une valeur supérieure à un seuil donné, on réalise une sorte de détection de contours, assez rudimentaire, mais suffisante pour mettre en évidence la difficulté du problème : d'une part, les points obtenus ne forment pas des courbes en général, et d'autre part, le résultat obtenu est très dépendant du seuil choisi (voir encadré).

La théorie du *scale-space*, née dans les années 1990, résoud partiellement le problème en prônant l'analyse multiéchelle d'une image à l'aide de lissages de degrés croissants. Cette théorie, formalisée de façon élégante au moyen d'équations aux dérivées partielles, laisse néanmoins entier le problème du réglage du seuil

¹David Marr, *Vision*, 1982

de détection, qui interdit toute détection purement automatique.

Courbes de niveaux

L'objet mathématique le plus simple pour décrire la géométrie d'une image (et donc les contours des objets qu'elle contient) sont ses courbes de niveaux, c'est-à-dire les courbes bordant les régions d'intensité supérieure à un seuil donné (ces courbes étant calculées pour chaque valeur possible du seuil). Cette décomposition est légitimée par le fait que, comme le remarqua Wertheimer au début du XXe siècle, notre perception visuelle est essentiellement indépendante de l'intensité absolue des scènes observées. Parmi ces courbes, les candidats à être des contours sont celles qui délimitent une forte variation d'intensité dans l'image. En calculant la probabilité qu'un tel contraste se produise par hasard le long d'une courbe, et en ne conservant que les courbes de niveaux pour lesquelles cette probabilité est inférieure à une probabilité d'erreur fixée à l'avance (détection *a contrario*), on s'affranchit dans une large mesure du problème du seuil tout en garantissant la détection de contours sous la forme de courbes (et non de pixels isolés).

Géométrie numérique

Analyser, comparer, ou mettre en correspondance des courbes extraites d'une image requiert alors des modèles mathématiques à la frontière entre géométrie différentielle, probabilités et analyse numérique. Le calcul effectif d'un invariant différentiel comme la courbure, dont la définition mathématique est pourtant très simple, soulève nombre de questions nouvelles et intéressantes liées à la discrétisation, à l'imprécision et à la structure multiéchelle des courbes observées. Ces dernières années ont vu se développer des équations aux dérivées partielles géométriques (et des schémas numériques associés) permettant de déformer des courbes à des fins de lissage ou de mise en correspondance.

Stéréovision

Dans ce même registre géométrique, un domaine de l'analyse d'images assez bien maîtrisé aujourd'hui est celui de la reconstruction du relief. Il s'agit de reproduire un mécanisme à l'oeuvre dans notre cerveau, la perception du relief à partir de la combinaison des images délivrées par chaque oeil. Cette technique sert à réaliser des cartes précises du relief de la terre (ou d'autres planètes, comme Mars récemment) à partir de couples d'images satellites.

Mathématiquement, on peut montrer qu'il suffit de mettre en correspondance sept points entre deux images pour identifier précisément la position relative des deux points de vue, et décrire la géométrie *épipolaire* associée, qui explicite la relation entre les deux projections d'un

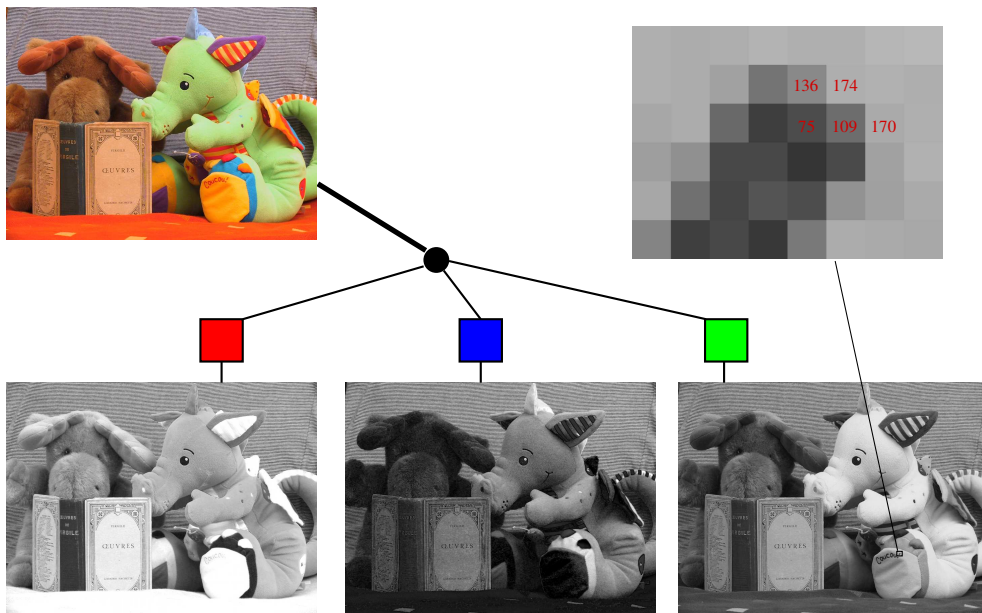
même point physique (voir encadré). Cette théorie se formalise de façon élégante et profonde dans le cadre plus général de la géométrie projective, c'est-à-dire la géométrie des droites de l'espace passant par un point donné (très exactement ici, les rayons lumineux qui forment l'image).

Expliquer par les mathématiques, illustrer par l'image

Les exemples que nous avons évoqués ici ne prétendent pas donner une idée exhaustive des champs mathématiques concernés par les images numériques², mais laissent néanmoins apparaître l'omniprésence, sous des formes très diverses, des mathématiques dans leur modélisation. Enjeu technologique, objet mathématique, l'image numérique est aussi un outil d'illustration, ce qui lui confère une place assez particulière au sein des mathématiques appliquées. L'enseignement de cette discipline, et tout particulièrement les séances de travaux pratiques sur ordinateur, offrent ainsi une double perspective ; celle d'expliquer un phénomène lié au traitement d'images par un théorème mathématique, et celle d'illustrer par l'image une théorie mathématique *a priori* abstraite.

²Nous n'avons pas mentionné, par exemple, la récente théorie des ondelettes, qui est au coeur de la nouvelle norme de compression d'images (JPEG 2000).

Détail d'une image numérique



Une image numérique couleur est le résultat de la superposition de 3 images monochromes fondamentales (rouge, vert et bleu), simples tableaux de nombres dont chaque case code l'intensité du pixel correspondant.

Théorie de Shannon et B-splines

Selon la théorie de Claude Shannon, une fonction de deux variables réelles $f(x, y)$ assez régulière (précisément : à bande limitée) peut être reconstruite à partir de ses échantillons discrets $f(k, l)$ (k, l entiers) grâce à la formule

$$f(x, y) = \sum_{(k, l) \in \mathbb{Z}^2} f(k, l) \frac{\sin \pi(x - k) \sin \pi(y - l)}{\pi^2(x - k)(y - l)}. \quad (R)$$

Cette formule peut être utilisée pour appliquer précisément un zoom ou une rotation à une image numérique. Elle peut être approximée de façon plus locale à l'aide des fonctions splines, définies par récurrence par

$$\beta^0(x) = \begin{cases} 1 & \text{si } |x| \leq \frac{1}{2}, \\ 0 & \text{sinon,} \end{cases} \quad \text{et} \quad \forall n \in \mathbb{N}, \quad \beta^{n+1}(x) = \int_{x-1/2}^{x+1/2} \beta^n(t) dt.$$

Une fois choisi l'entier n (l'ordre d'approximation), l'équation (R) devient

$$f(x, y) \simeq \sum_{(k, l) \in \mathbb{Z}^2} c(k, l) \beta^n(x - k) \beta^n(y - l),$$

les coefficients $c(k, l)$ étant calculés de telle sorte que la formule soit exacte pour x et y entiers. Quand n tend vers l'infini, cette approximation rejoint la formule de reconstruction de Shannon (R).

Réduction d'une image



Pour réduire (ici, d'un facteur 4) la taille d'une image numérique en conservant le cadrage initial, extraire un point sur 16 de l'image initiale (encadré précédent) est une technique particulièrement maladroite, qui produit un phénomène d'*aliasing* (image de gauche) : effets de moiré sur les motifs réguliers (ici, le fond rayé et le dos du livre), perte de connexité des lignes (caractères illisibles), crénelage des contours, etc. Une technique sensiblement meilleure (mais néanmoins non optimale) consiste à moyenner les valeurs présentes dans chaque carré 4×4 de l'image initiale pour construire le nouveau pixel.

Déflouage d'une image



Améliorer la qualité d'une image floue (à gauche) est un exemple de problème inverse : toute technique de résolution "directe" est instable et conduit à une amplification inacceptable du bruit de mesure (au centre). On peut remédier à cela en cherchant des solutions approchées mais plus régulières (à droite, régularisation par variation totale).

Détection de contours



La plupart des techniques d'analyse automatique d'images commencent en général par extraire des contours (transitions fortement contrastées entre un objet et le fond). En sélectionnant simplement les points de l'image de gauche dont le contraste dépasse un certain seuil (au centre), on met en évidence 3 difficultés essentielles : l'automatisation (le résultat est ici très sensible au seuil choisi), l'élimination des contours parasites, et le regroupement des points de contours en courbes (non réalisé par cette méthode). Une technique récente, effectuant une sélection probabiliste des courbes de niveaux bien contrastées de l'image, répond assez efficacement à ces problèmes et produit un résultat plus exploitable (à droite).

La géométrie épipolaire

Si un même point physique se projette dans deux images en des points (x_1, y_1) (image 1) et (x_2, y_2) (image 2), alors

$$(x_2 \ y_2 \ 1) F^t (x_1 \ y_1 \ 1) = 0,$$

où F est une matrice 3×3 de rang 2 (définie à une constante près), qui ne dépend que de la position relative des deux points de vue. L'équation ci-dessus régit la géométrie épipolaire de la scène : elle signifie que le point (x_2, y_2) appartient à une certaine droite (appelée droite épipolaire) définie par le point (x_1, y_1) . Cette droite correspond à la projection du trajet optique de l'image 1 sur le plan de l'image 2. Les droites épipolaires ont toutes un point commun, l'épipôle, projection du centre optique associé à l'image 1.