

L'IMAGERIE NUMÉRIQUE

Pour réaliser leurs travaux¹, les frères Chudnovsky ont fait appel à diverses méthodes issues du domaine de l'imagerie numérique. Ces différentes méthodes renvoient à trois grandes étapes : (1) identifier des points caractéristiques, (2) faire l'appariement de ces points entre les différentes photos de la tapisserie, et (3) effectuer une transformation géométrique de chaque photo pour aligner les points caractéristiques et reconstituer la tapisserie complète.

Jean Meunier
Université de Montréal

Identifier des points caractéristiques

En traitement d'images on a recours fréquemment à l'extraction de points caractéristiques (ou points d'intérêt) pour diverses applications comme la détection du mouvement, le recalage d'images, l'assemblage de photos, la production d'un panorama, la reconnaissance d'objets, etc. Il s'agit de points bien définis et stables dans l'image. Par exemple, on peut chercher à détecter des « coins » en cherchant dans l'image les points où il y a un changement brusque dans la direction des contours des objets comme pour les quatre coins d'un objet rectangulaire. Cette approche permet d'obtenir automatiquement un ensemble de points caractéristiques sur chaque photo. La figure ci-contre donne un exemple² de points caractéristiques (en vert) obtenus avec le détecteur SURF (Speeded Up Robust Features³) qui est populaire aujourd'hui en imagerie numérique.



Apparier des points correspondants

Lors de la prise des photos de la tapisserie, on a pris soin de s'assurer que celles-ci se recouvraient suffisamment pour pouvoir les aligner plus tard. Cette zone de recouvrement est relativement importante (de l'ordre de 50%) et permet de trouver les points caractéristiques *communs* de photos de tapisserie qu'on souhaite aligner. Pour cela, on analyse une petite région (appelée voisinage) autour de chaque point caractéristique et on fait l'appariement des points qui possèdent un voisinage similaire. Pour mesurer la similarité entre deux voisinages a et b on peut par exemple calculer la somme des différences (en valeur absolue) de couleur de chaque pixel :

$$\sum_i \sum_j |a_{ij} - b_{ij}|,$$

où i et j représentent la position (ligne, colonne) du pixel dans le voisinage. Cette mesure sera petite si la similarité est grande. Cependant cela peut poser problème si l'illumination varie entre la prise de photos puisque deux régions identiques apparaîtront alors différentes en intensité dans ce cas. On peut alors miser sur une mesure plus robuste : l'intercorrélation, *Corr*.

1. Voir dans ce numéro La quête de la licorne, p. 14-15.
2. Exemple adapté de <http://www.wolfram.com/services/education/seminars/s51.html>
3. Voir http://fr.wikipedia.org/wiki/Speeded_Up_Robust_Features

On la définit comme suit :

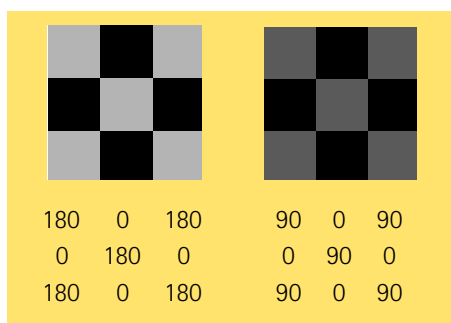
$$\text{Corr}(a,b) = \frac{\text{Cov}(a,b)}{\sqrt{\text{Cov}(a,a) \cdot \text{Cov}(b,b)}}$$

où $\text{Cov}(a,b) = \sum_i \sum_j (a_{ij} - \bar{a}) \cdot (b_{ij} - \bar{b})$

est appelée covariance. \bar{a} et \bar{b} sont les moyennes respectives des a_{ij} et b_{ij} .

Elle atteint une valeur maximale de 1 lorsque les voisinages a et b sont identiques à une constante multiplicative ou additive près.

Prenons un exemple simplifié où seule l'intensité (ou luminance) du pixel est utilisée⁴. Considérons des voisinages de 3×3 pixels (ils sont plus grands en pratique) pour a et b . Ces voisinages sont identiques sauf les valeurs d'intensité de chaque pixel : celles de b valent la moitié de celles de a (comme cela pourrait être le cas avec un changement d'éclairage par exemple).



La somme des valeurs absolues des différences nous donne évidemment une différence importante, égale à $(180-90) \times 5 = 450$. Par contre, pour l'inter-corrélation, on a : $\bar{a} = 100$ et $\bar{b} = 50$, puis

$$\begin{aligned} \text{Cov}(a,a) &= 5 \cdot (180 - 100)^2 + 4(0 - 100)^2 = 72\,000, \\ \text{Cov}(b,b) &= 5 \cdot (90 - 50)^2 + 4(0 - 50)^2 = 18\,000, \\ \text{Cov}(a,b) &= 5 \cdot (180 - 100) \cdot (90 - 50) + 4(0 - 100) \cdot (0 - 50) \\ &= 36\,000 \\ \text{Corr}(a,b) &= \frac{36\,000}{\sqrt{72\,000 \cdot 18\,000}} = 1. \end{aligned}$$

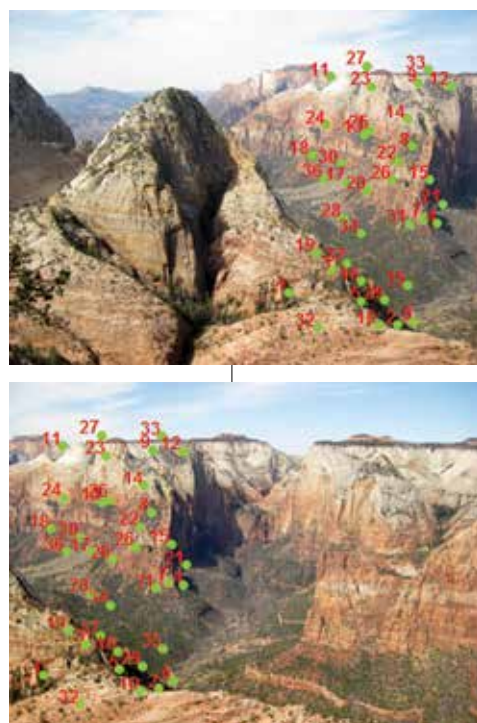
On a donc une corrélation maximale comme souhaitée car les deux voisinages sont identiques sauf pour l'illumination. En pratique, l'intercorrélation ne sera pas égale

4. Une image couleur aurait 3 composantes R, V et B. Voir « Les images sur la toile un défi de taille », *Accromath 7.2, 2012*

à 1 mais plutôt proche de 1 pour les régions suffisamment similaires qui seront mises en correspondance.

Pour des voisinages plus grands et de nombreux points caractéristiques à comparer, cette procédure peut être coûteuse en temps de calcul. De plus, pour tenir compte des déformations de la tapisserie d'une photo à la suivante, les frères Chudnovsky ont proposé de comparer chaque voisinage a à une transformation affine de b (rotation, translation, mise à l'échelle). L'utilisation d'un superordinateur parallèle s'est avérée très utile ici, car les différents processeurs ont pu se partager la tâche en travaillant chacun sur une partie de la tapisserie.

Les images ci-haut illustrent cette étape où chaque point caractéristique d'une image et le point correspondant dans l'autre image sont identifiés par un numéro (en rouge). La méthode utilisée ici est différente (SURF) mais le principe demeure le même.



Transformer pour réunir

L'exemple du panorama

Lorsqu'on souhaite produire un panorama avec plusieurs photographies, on a recours à un assemblage qui consiste à combiner plusieurs photos se recouvrant partiellement. Cependant il faut en général appliquer une transformation géométrique (projection perspective) aux images afin de bien les aligner lors de la construction du panorama. Cette transformation peut être obtenue automatiquement

en cherchant à aligner le mieux possible les points caractéristiques des deux photos dans leur zone de chevauchement.



La figure ci-haut illustre ce principe. La seconde photo a été transformée par projection perspective pour obtenir une superposition adéquate des points caractéristiques et placée sous la première photo pour mieux voir la ligne de démarcation entre les deux images. Cette ligne peut être éliminée en effectuant un fondu approprié entre les deux images.

Pour un panorama classique le type de transformation est connu (projection perspective) ; par contre pour la tapisserie, la transformation peut être beaucoup plus complexe et une simple transformation (perspective ou autre) ne suffit pas. Les frères Chudnovsky ont sans doute eu recours à une autre technique bien connue en infographie qu'on appelle le warping.

Le warping

Le warping consiste à transformer de façon naturelle, graduelle et fluide une image initiale en une image finale différente en alignant deux ensembles de points caractéristiques tout en permettant des transformations complexes et réalistes. Pour cela on utilise souvent la triangulation de Delaunay qui consiste en un ensemble de triangles couvrant une image de façon optimale et dont les sommets sont les points caractéristiques. Les images ci-contre illustrent les principes sous-tendant cette technique. Dans cet exemple⁵, la triangulation permet d'effectuer le warping de l'image du haut vers l'image du milieu. Un résultat intermédiaire est donné en bas en effectuant un fondu des couleurs

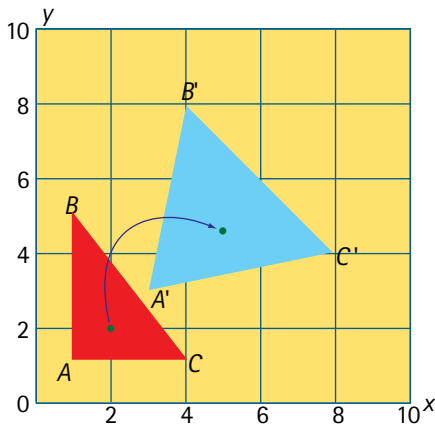


entre les deux images; on parle alors de morphing entre les deux images. Chaque triangle de l'image de gauche a un triangle correspondant dans l'image du centre afin de calculer les transformations nécessaires. Dans la reconstitution de la tapisserie, on utilise cette méthodologie pour aligner les photographies entre elles dans leurs régions de chevauchement.

On utilise donc l'emplacement des points caractéristiques obtenus précédemment pour construire une triangulation de Delaunay dans la zone de chevauchement des photographies. Ensuite, connaissant le déplacement des points caractéristiques entre deux photos (puisque l'appariement des points est connu), il ne reste qu'à interpoler le déplacement des autres pixels à l'intérieur de chaque triangle à l'aide d'une simple transformation linéaire. La transformation complexe des photos est donc réduite à un ensemble de transformations beaucoup plus simples (linéaires) pour chacun des triangles.

5. Exemple adapté de <http://www.mukimuki.fr/flashblog/labs/warping/MorphApp.html>

La figure suivante illustre ce principe à l'aide d'un exemple simplifié. Le triangle rouge doit subir une transformation afin de le superposer sur le triangle bleu comme pour les paires de triangles utilisés dans le warping. On connaît la position des 3 points caractéristiques (sommets) pour les deux triangles mais qu'en est-il du point (2, 2) situé à l'intérieur du triangle rouge (point vert)? Où ce point se retrouvera-t-il dans le triangle bleu?



Pour répondre à cette question on doit d'abord trouver la transformation affine (linéaire) entre les deux triangles, c'est-à-dire une transformation dont la formulation est la suivante :

$$\begin{aligned}x' &= ax + by + c \\y' &= dx + ey + f\end{aligned}$$

où (x, y) est la position initiale des pixels dans le triangle rouge et (x', y') la position finale des pixels dans le triangle bleu. Afin de déterminer les paramètres a, b, c, d, e et f de cette transformation, il suffit de miser sur le déplacement connu des 3 sommets de ces deux triangles afin d'obtenir un petit système d'équations en remplaçant dans la transformation affine précédente les valeurs (x, y) et (x', y') connues des sommets.

Ainsi, la transformation de $A(1, 1)$ vers $A'(3, 3)$ nous donne :

$$\begin{aligned}a \times 1 + b \times 1 + c &= 3 \\d \times 1 + e \times 1 + f &= 3\end{aligned}$$

Celle de $B(1, 5)$ vers $B'(4, 8)$:

$$\begin{aligned}a \times 1 + b \times 5 + c &= 4 \\d \times 1 + e \times 5 + f &= 8\end{aligned}$$

Et celle de $C(4, 1)$ vers $C'(8, 4)$:

$$\begin{aligned}a \times 4 + b \times 1 + c &= 8 \\d \times 4 + e \times 1 + f &= 4\end{aligned}$$

Ce système linéaire de 6 équations à 6 inconnues est en fait constitué de deux petits systèmes indépendants de 3 équations à 3 inconnues (a, b, c , d'une part; et d, e, f , d'autre part). Il se résout aisément pour donner :

$$\begin{aligned}a &= \frac{5}{3}, \quad b = \frac{1}{4}, \quad c = \frac{13}{12}, \\d &= \frac{1}{3}, \quad e = \frac{5}{4}, \quad f = \frac{17}{12}.\end{aligned}$$

On peut maintenant répondre à la question : que devient le point (2, 2) du triangle rouge? En substituant dans la transformation affine la valeur des paramètres et la position $(x, y) = (2, 2)$ on obtient

$$\begin{aligned}x' &= \frac{5}{3}x + \frac{1}{4}y + \frac{13}{12} \approx 4,92, \\y' &= \frac{1}{3}x + \frac{5}{4}y + \frac{17}{12} \approx 4,58.\end{aligned}$$

C'est ce principe qui permet d'interpoler le déplacement de tous les pixels à l'intérieur de chaque triangle couvrant la zone de chevauchement de la tapisserie et de pouvoir les recaler (aligner). Notons que les coordonnées finales $(x', y') = (4,92; 4,58)$ ne sont pas des entiers ; il faudra donc recourir à une autre forme d'interpolation pour distribuer la couleur du pixel (2,2) sur les pixels voisins de coordonnées entières dans l'image finale. Au besoin il restera aussi à faire un ajustement de l'intensité de certains pixels pour tenir compte de changements d'illumination entre la prise de deux photos en utilisant par exemple les changements observés au niveau des points caractéristiques et en interpolant ces valeurs à l'intérieur de chaque triangle.