

Suivi dans une séquence d'images

Elie MICHEL
Lycée Saint-Louis,
MP* 2

16 novembre 2012

Introduction

Ce texte a pour but de synthétiser les recherches que j'ai menées dans le cadre du TIPE (Travail d'Intérêt Personnel Encadré) pour les concours des écoles d'ingénieurs de 2013. Le thème imposé était *Similitudes et invariances* et après avoir fait des recherches dans des domaines très diverses mon choix s'est porté sur le suivi dans une séquence d'images. Le but est de reconnaître un même objet ou motif d'une image à une autre. Ses applications sont nombreuses et très diversifiées puisqu'il est utilisé aussi bien par les particuliers, avec des programmes de génération de panoramiques à partir de plusieurs photographies se superposant par exemple, ou dans le cinéma, lors de la superposition d'images synthétiques à des images réelles ainsi que dans la robotique.

Nous verrons dans un premier temps comment reconnaître un même motif d'une image à une autre à l'aide de repères insensible aux différentes modifications que peut subir cet objet, comme un changement de taille, une rotation, une déformation ou un changement de luminosité. Nous verrons ensuite des applications pratiques de ce suivi avec notamment le match moving, c'est-à-dire l'extraction d'une séquence vidéo des mouvements de la caméra.

1 Tracking

Dans toute cette partie, on considère une image en niveaux de gris (correspondant aux niveaux de luminosité) mais le cas d'une image en couleur est en fait identique puisque celle-ci n'est qu'un triplet de trois passes représentant les niveaux de luminosités de rouge, de vert et de bleu.

1.1 Idée générale

Le traitement d'image serait une tâche fastidieuse si l'on devait l'effectuer manuellement et son automatisation permet de plus des applications bien plus précises et rapides sur des séquences d'images parfois longues. Le problème est de trouver un moyen de reconnaître de façon automatique un même objet sur deux images différentes. Nous allons pour cela repérer des points remarquables de l'image et les faire correspondre avec ceux de l'autre image. Cette méthode permet d'avoir, une fois la correspondance entre les deux images faite, directement le déplacement de l'objet entre les deux images, contrairement aux méthodes qui utiliseraient une analyse de l'image dans sa globalité.

Un point remarquable est une singularité de l'image permettant de la reconnaître, de la caractériser. La question est de savoir comment choisir ces points. Ce choix est très important puisque ces points ne doivent comporter ni trop d'information, ni pas assez. S'ils contiennent trop d'informations, il y a un risque de surapprentissage, c'est-à-dire que les données sur l'objet seront trop spécifiques et qu'il ne sera alors reconnu que dans des conditions très précises. A l'inverse, s'ils n'en contiennent pas assez, il y aura de grandes possibilités de confusion.

1.2 Choix des points remarquables

1.2.1 Invariance

La solution est de choisir des points invariants, de caractériser l'objet par des données qui ne dépendent pas de la façon de la regarder. Nos points remarquables doivent donc être le moins sensibles possible aux :

- changements d'échelle ;
- rotations ;
- changements d'éclairage ;
- déformations dues à la perspective.

L'invariance par changement de perspective n'est pas réalisable à la perfection puisque reconnaître un objet de profile lorsqu'on ne l'a vu que de face est a priori impossible mais pour de petites variations de la perspective, les invariances précédentes permettent un certaine robustesse dans le sens où ces déformations correspondent localement à des rotations et changement d'échelle. La difficulté sera de prendre en compte la déformation globale lors de la correspondance des points entre les deux images.

Remarque On ne cherche pas l'invariance par translation des points remarquables puisque ce sont justement ces modifications que l'on cherchera par la suite à exploiter pour en déduire des mouvements par exemple.

Il y a deux aspects à prendre en compte dans ces invariances.

- Elles doivent permettre de repérer les mêmes points d'une image à une autre.
- Mais elle doivent également permettre de reconnaître ces points. Ils doivent donc avoir des caractéristiques qui suivent ces invariants.

1.2.2 Sélection

On s'intéresse ici à la méthode de sélection des points remarquables. L'invariance par changement de luminosité impose d'utiliser pour cela des grandeurs relatives, des variations, et non la valeur absolue de l'image en un point. De plus, l'invariance par rotation implique que le critère de sélection soit isotrope. La recherche des extrema répond à ces deux critères et est a priori également invariante par changement d'échelle. Mais à priori seulement. En effet, l'image est en pratique représentée de façon discrète. La taille des échantillons joue donc un rôle.

2 Matching

Une fois une image caractérisée par un ensemble de points de repère, comparer deux images revient à comparer leurs ensembles de points de repère. On cherche ainsi à trouver des correspondances entre ces points et il faut pour cela trouver dans une image le point ressemblant le plus à un point donné de l'autre image. Ce problème est un problème très souvent rencontré en intelligence artificielle et en particulier en apprentissage automatique : la recherche du plus proche voisin.

2.1 Plus proche voisin

Le problème de recherche du plus proche voisin d'un point dans l'ensemble des points de repère d'une image peut plus généralement être formalisé comme suit :

Problème

Soit E un espace euclidien de dimension d
 Soit $V = (v_k)_{k \in \{1, \dots, n\}}$ une famille de n points de E
 Soit $u \in E$
 Comment trouver le point v_k de V telle que $\|u - v_k\| = d(u, V)$?

Remarque Ici on munit $E = \mathbb{R}^d$ est muni du produit scalaire usuel $(x, y) \mapsto \sum x_i y_i$.

Une première approche naïve consiste à tester toutes les possibilités. On doit donc calculer n normes et les comparer. En évitant de

3 Applications

Après avoir vu comment retrouver un objet ou un motif dans différentes images, il est légitime de se demander pourquoi de telles méthodes sont intéressantes; dans quel cas elle sont utilisées. Les applications du suivi de formes sont multiples et présentes dans des domaines très divers. Citons par exemple :

- La création d'image panoramique par recollement d'images se chevauchant ;
- La stabilisation vidéo, que ce soit pendant la prise de vue (au niveau de l'appareil photo, pratique lorsque le zoom est important) ou a posteriori en rognant l'image pour limiter les tremblements.
- L'identification de lieux (pour en connaître le nom à partir d'une photo) ou de personnes (ex : Picasa, Facebook tentent de reconnaître les personnes présentes sur les photos)
- L'identification d'empreintes digitales (ou autres comparaisons précises et fastidieuses)
- Le match moving
- La réalité augmentée (temps réel)
- scanning 3d (structure from motion)
- Motion capture
- domaine médical ?
- surveillance, radar routiers
- numérisation de texte

3.1 En deux dimensions

Panoramique On s'intéresse ici plus particulièrement au problème du panoramique. Une première étape permet de placer globalement les images les unes par rapport aux autres. Pour cela, on sélectionne des points de repère aléatoirement dans la liste et on cherche le point le plus proche. Une régression linéaire permet d'obtenir une première estimation de la position relative des images. On entreprend dans un second temps de déformer les images de sorte qu'elles correspondent en tout point en les comparant localement.

Motion capture Il est important dans ce cas de bien localiser les repères dans les premières images (on analyse tout de même plusieurs images au cas où la première (ou son analyse) comporterait des imperfections) puis ensuite, sachant que les images sont peu espacées, il suffit de se concentrer sur un voisinage de la position du repère dans l'image précédente. Cette méthode peut être utilisée pour tout suivi dans une vidéo puisque les images sont très peu espacées temporairement.

4 [Debug] Pistes de recherche

Filtrage anisotropique J'ai tenté une rapide implémentation du filtre de Perona Malik mais j'ai dû manquer quelque chose car il y a des points « brûlés » sans raison, notamment au niveau des angles. On retrouve cependant bien un effet de flou pour $c = 1$ dans la formule d'évolution :

$$\frac{\partial I}{\partial t}(x, y, t) = \vec{\nabla} \cdot (c(|\vec{\nabla}(I)|) \cdot \vec{\nabla}(I))$$

C'est-à-dire pour $c = 1$:

$$\frac{\partial I}{\partial t}(x, y, t) = \vec{\nabla}^2(I)$$

ce qui est vérifié pour une convolution par une gaussienne d'écart type $2\sqrt{t}$.