

Suivi d'objets dans une séquence d'images

Elie MICHEL
Lycée Saint-Louis,
MP* 2

15 mai 2013

Introduction

On ne pense pas à la façon dont on reconnaît des objets mais c'est pourtant un phénomène complexe. Il se pose la question de sa reproduction par une machine dans de nombreuses situations. Les applications de la reconnaissance d'objets 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 ou motion tracking (calcul du mouvement d'une caméra à partir des images capturées)
- La réalité augmentée (équivalent en temps réel du match moving)
- Le scanning 3D et photogramétrie (Reconstruction d'une géométrie 3D à partir d'une vidéo)
- Le motion capture (capture des mouvements d'un acteur dans le but de les faire reproduire à un personnage virtuel)
- La surveillance (radar routiers)

Toutes ces tâches nécessitent un traitement long et précis d'une grande quantité d'images qui nécessiterait à un être humain un temps considérable. Or, il n'est pas toujours possible d'attendre, notamment en ce qui concerne les applications en temps réel comme la réalité augmentée ou encore les radars automatiques. Il est donc nécessaire d'automatiser ces tâches au maximum.

L'enjeu de ce TIPE est donc de comprendre comment un ordinateur peut extraire d'une séquence d'image des informations diverses. Mais avant de s'intéresser à l'ensemble de ces images, nous devons trouver un moyen de comprendre ces images individuellement afin de pouvoir ensuite les comparer. La première partie traite donc de l'extraction de primitives d'une image afin de la caractériser par des grandeurs que l'on peut comparer. Dans un deuxième temps, nous commencerons par comparer

seulement deux images à l'aide des primitives précédemment construites et enfin nous verrons comment il est possible à partir de comparaisons d'images de sortir différentes informations comme la déplacement relatif de deux images dans la plan ou même dans l'espace.

Bien que l'on puisse imaginer de très nombreuses façons de représenter une image en informatique, on manipulera des images sous forme matricielle, c'est-à-dire sous la forme d'un tableau de pixel. C'est en effet la forme brute d'image la plus commune, et même si pour des besoins de compression, notamment en ce qui concerne les vidéo, la description des images n'est pas toujours celle-ci, le traitement d'image se fait généralement à partir des données décompressées.

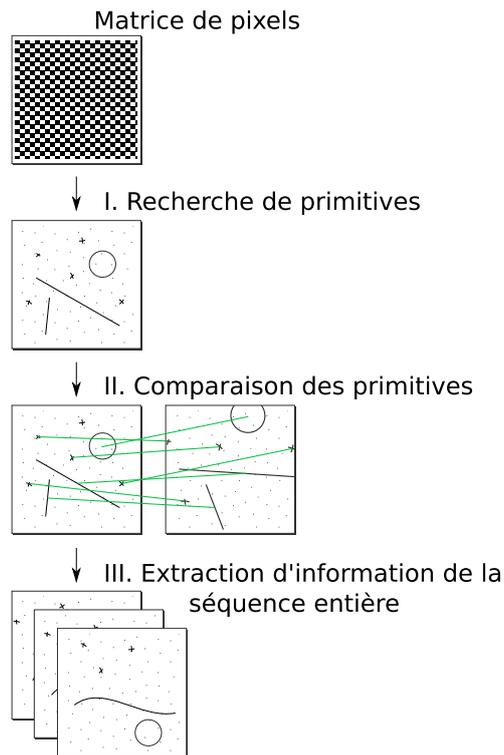


FIGURE 1 – Plan du TIPE

1 Recherche de primitives

1.1 Nécessité de primitives caractérisant les objets

Afin de reconnaître un objet ou une forme donnée, il faut le caractériser par des données simples à manipuler et à **comparer**. La possibilité de comparer des **repères** est capitale. En effet, il serait illusoire de considérer que la représentation d'un objet est exactement la même d'une image à l'autre. Il est donc nécessaire de disposer d'une mesure de similitude permettant de formaliser le fait de "ressembler" à une image. Par conséquent, les repères que l'on cherche doivent être la moins dépendant possible à la représentation des objets.

Si l'on veut traiter une séquence d'images, il faut d'une part trouver les mêmes repères d'une image sur l'autre et d'autre part être en mesure de reconnaître ces repères. On a donc besoin de méthodes

robustes à la fois pour la **sélection** et pour la **caractérisation** des repères. On parle aussi parfois de détection et description.

La représentation des objets peut être altérée par différents facteurs regroupés en deux catégories :

- Une modification de l'objet et de sa situation :
 - Changement de l'éclairage
 - Changement de l'angle de vue obstruant certaines parties de l'objet
- Une modification de la projection de l'objet :
 - Rotation
 - Changement d'échelle
 - Déplacement

D'autres types de perturbations peuvent intervenir comme des déformations des objets, des distorsions de l'image dues à un apllanétisme imparfait des lentilles du système de capture, reflets, etc. Nous ne les traiterons cependant pas dans ce document.

Les secondes perturbations sont dues à l'aspect discret et orienté de la capture des images. En effet, lorsque l'on représente le plan de projection des images dans une matrice, il faut non seulement discrétiser ce plan, ce qui implique une échelle privilégiée mais aussi en sélectionner une partie bornée (position privilégiée) et une base indiquant l'orientation de la grille de pixels, ce qui particularise certaines directions. (Voir figures 2 et 3) Selection et caractérisation des repères doivent donc s'affranchir au maximum de ces particularités imposées uniquement par la représentation des objets et non par les objets eux-mêmes.

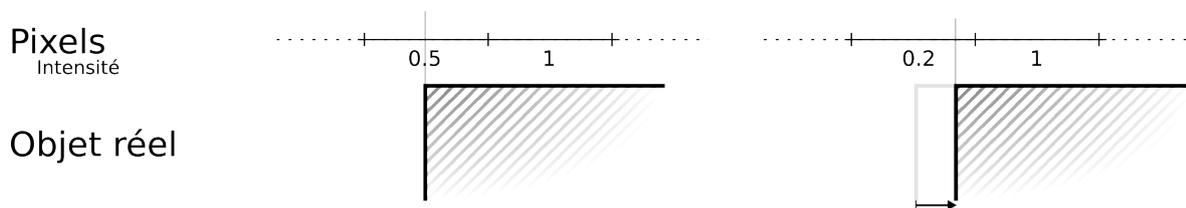


FIGURE 2 – À cause de la discrétisation, une translation de l'objet réel ne se traduit pas sur la grille de pixels par une translation

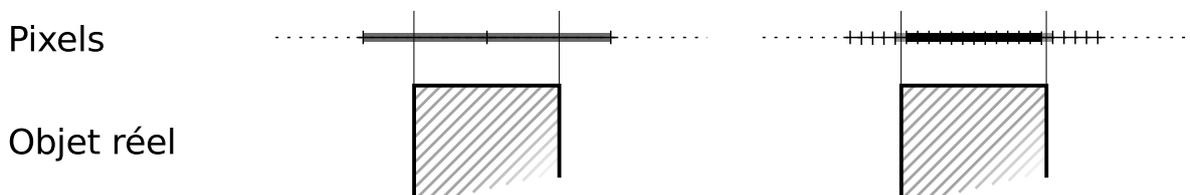


FIGURE 3 – Pour des objets de taille comparable à celle des pixels, la taille de leur image n'est pas exactement leur taille réelle

D'autre part, l'information portée par les repères doit être suffisamment précise mais pas trop. On cherche à enregistrer les propriétés propres à l'objet et non aux conditions qui l'entourent. Si l'information extraite est trop précise, il y a un risque de *surapprentissage* : les données sont trop spécifiques et la reconnaissance ne se fait alors que dans des conditions données, ce qui n'est pas ce que l'on cherche. A contrario, une information trop vague peut entraîner des confusions.

1.2 Problématique générale de la recherche d'invariants

Avant de nous intéresser aux méthodes utilisées en pratique pour trouver des repères, nous nous intéressons ici à ce qu'imposent les différentes invariances recherchées.

1.2.1 Luminosité

L'invariance par changement de luminosité nécessite l'exploitation de **grandeurs relatives**. Dans l'absolu, la valeur des pixels n'a aucune réalité physique : c'est juste une échelle de mesure dont on ne connaît pas la graduation exacte puisque celle-ci dépend de facteurs très variables comme le temps d'ouverture de l'obstruteur, la sensibilité du capteur ou le format de fichier utilisé. On doit donc étudier les **variations relatives** de ces valeurs, que ce soit entre les pixels ou au sein d'un même pixel entre les couleurs qui le composent.

Remarque En général, on traite des images en niveau de gris quitte à séparer les différents canaux de couleur pour les traiter indépendamment. Il y a par contre plusieurs façon de séparer les couleurs (Rouge-Vert-Bleu, Teinte-Saturation-Luminosité, etc).

1.2.2 Rotation

L'invariance par rotation nous oblige à utiliser des critères de sélection isotrope. Il faut donc traiter au maximum toutes les directions de la même façon et non privilégier les directions horizontale et verticale dans l'image qui sont des artefacts dûs à la forme matricielle des images que l'on traite.

Utiliser des critères de sélection isotrope ne signifie cependant pas s'interdire d'exploiter l'orientation dans les caractéristiques d'un repères pourvu que celle-ci ne soit pas basée sur la direction de capture. On peut par exemple indiquer la direction de plus forte pente de l'image en un point. L'utilisation de caractéristiques d'orientation est même utile pour trouver la rotation entre deux images.

1.2.3 Changement d'échelle

L'invariance par changement d'échelle est plus complexe à gérer. En effet, l'échantillonnage effectué lors de la capture rend les détails trop petits invisibles (il y a une perte d'information) mais rend aussi les objets trop grands méconnaissables à cause de l'apparition de détails et de bruit et à cause de la limitation de la taille du cadre qui peut couper ces objets. Autant les détails plus fins que la taille des pixels ne peuvent être retrouvés, autant il est possible de donner aux objets plus grands l'apparence qu'ils auraient s'ils avaient été capturés avec une moins bonne fréquence d'échantillonnage. Pour cela, on peut simuler la capture d'une image par un système moins précis.