

IFT 6145
Vision tridimensionnelle
Estimation du mouvement I

Sébastien Roy
Département d'informatique et de recherche opérationnelle
Université de Montréal

hiver 2005

En bref...

Chapitre 8 du livre *Trucco & Verri*.

- Séquences d'images.
- Mouvement, flux optique
- ...

Séquence d'images

Séquence d'images

Une série de N images, acquises successivement à des instants séparés par un interval de temps Δt .

Notes

- Δt doit être petit (discrétisation temporelle dense).
- environ 24 à 60 images par secondes.
- Déplacement observé entre deux images successives < 2 pixels.

Séquences d'images

En général

Si on suppose une **illumination constante** et des **objets mats**, les variations observées de l'images sont causées **uniquement** par les mouvements des objets ou de la caméra (ou une combinaison des deux).

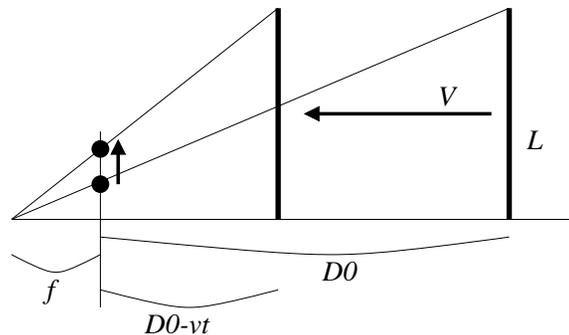
- La dimension *temporelle* ajoute de l'information pour analyser la scène 3D.
- Permet d'extraire des propriétés de la scènes sans avoir a faire une reconstruction complète et sans calibration.

Exemple d'application : Temps d'impact

Problème

Soit un objet rigide qui se dirige vers la caméra. Quel est le temps τ qui reste avant l'impact ?

On peut trouver le temps sans connaître la taille de l'objet ou même sa vitesse. On utilise seulement l'image.



Soit un objet de taille L se déplaçant à la vitesse V et placé à la position D_0 au temps $t = 0$ et à la position $D = D_0 - Vt$ au temps t .

Temps d'impact

Note

Les valeurs de L , D_0 , V , f (distance focale), sont inconnues.

Le temps d'impact peut être exprimé comme

$$\tau = \frac{D}{V}$$

La taille apparente de l'objet est $l(t)$ et est donnée par

$$l(t) = f \frac{L}{D}$$

Temps d'impact

La dérivée temporelle de $l(t)$ est

$$l'(t) = \frac{dl(t)}{dt} = -f \frac{L}{D^2} \frac{dD}{dt} = f \frac{LV}{D^2}$$

Le ratio entre $l(t)$ et $l'(t)$ est

$$\frac{l(t)}{l'(t)} = \tau$$

En mesurant $l(t)$ et $l'(t)$ dans l'image, on obtient le temps d'impact τ .

L'analyse du mouvement

Problèmes classiques en analyse du mouvement :

Correspondance

Quel élément de la prochaine image correspond à un élément donné de l'image ?

Reconstruction

Si on dispose de points mis en correspondance, quel sont leur positions et mouvements 3D ?

Segmentation du mouvement

Si on est en présence d'objets au mouvements indépendants, isoler ces objets, et trouver leur positions et mouvements 3D respectifs.

Liens avec la stéréoscopie

- Les images contiennent de faibles déplacements, donc les disparités sont très faibles.
- La scène n'est pas nécessairement rigide. Les objets peuvent se déplacer indépendamment ou être non-rigides.

Exemples :

Scène rigide (seule la caméra se déplace)

Caméra sur une auto ou robot qui se déplace.

Objets rigides multiples

Caméra qui surveille le trafic routier à une intersection.

Un ou plusieurs objets non-rigides

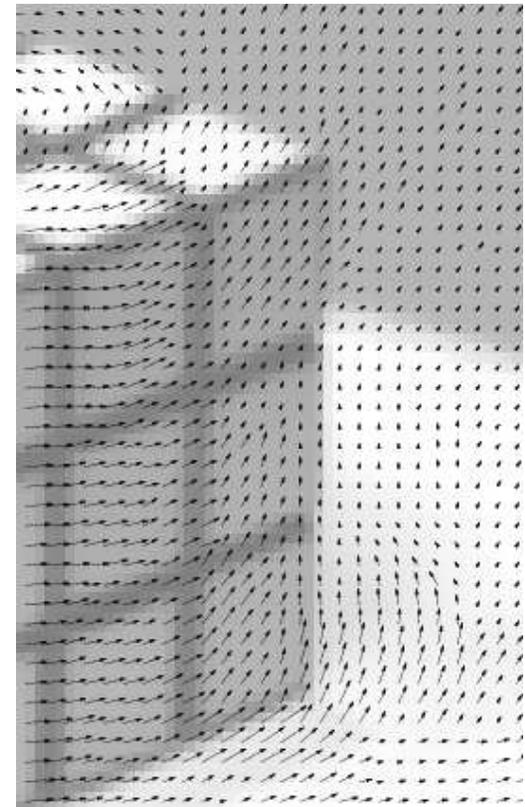
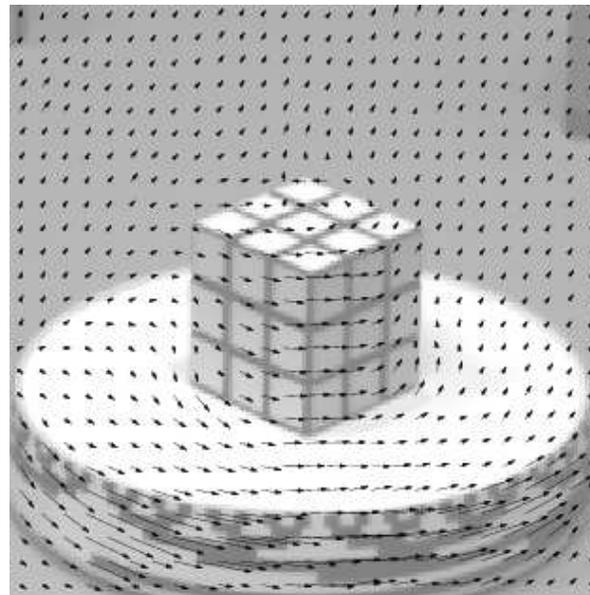
Cellules vivantes observées par un microscope.

Angiographie (cœur sous rayons-X)

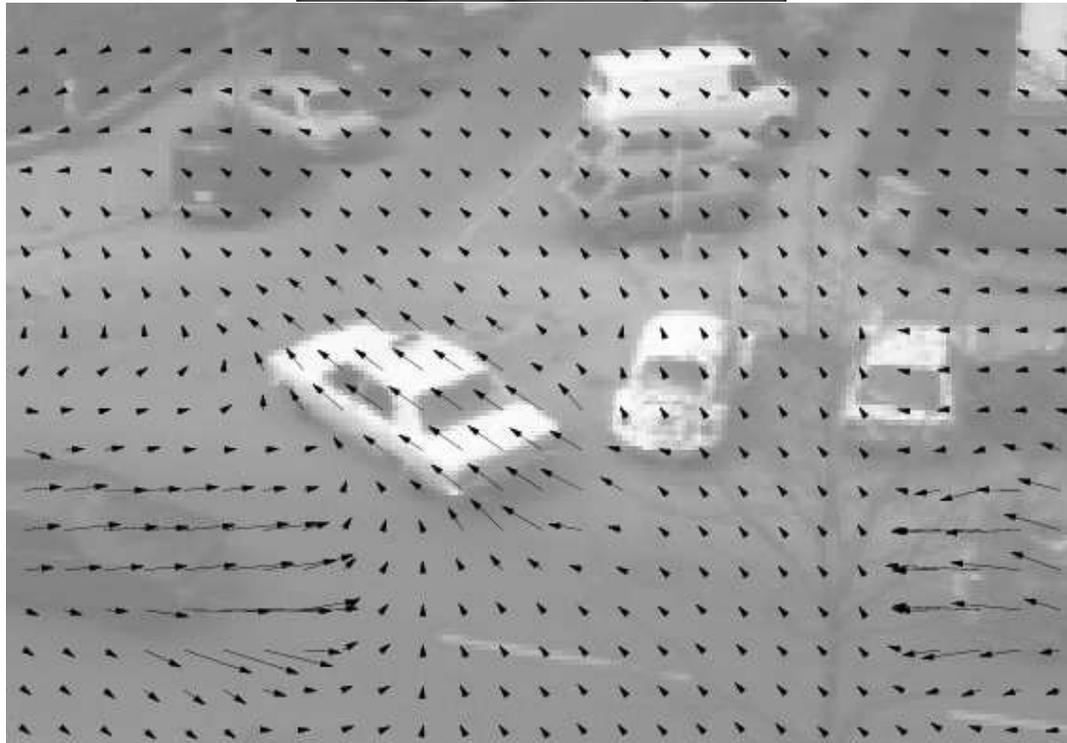
Flux optique

Flux optique

Champs de vitesse décrivant le mouvement apparent des motifs d'intensités de l'image.



Flux optique



Correspondance

Résoudre la correspondance équivaut à estimer le flux optique. Trois approches :

Approche différentielle

Utilise les dérivées spatio-temporelles de l'image comme source d'information sur le mouvement.

Approche corrélation

Utilise la recherche d'un point correspondant par la corrélation entre les images.

Approche domaine fréquence

Utilise les changements en fréquence et en phase pour déterminer le mouvement.

Hypothèses

Les hypothèses changent selon le type de problème.

Pas trop restrictif

Objets non-rigides avec mouvements arbitraires

→ contrainte de lissage

Moyen

Objets rigides avec mouvements arbitraires

→ contrainte de lissage + un peu paramétrique

Moyen

Une scène rigides avec mouvement arbitraire de caméra

→ contrainte de lissage + paramétrique

Très restrictif

Un seul plan au mouvement arbitraire

→ contrainte paramétrique

Flux optique d'un objet rigide

Objet rigide + mouvement de caméra

Note

Un objet rigide en mouvement + caméra fixe

=

Un objet rigide fixe + caméra en mouvement

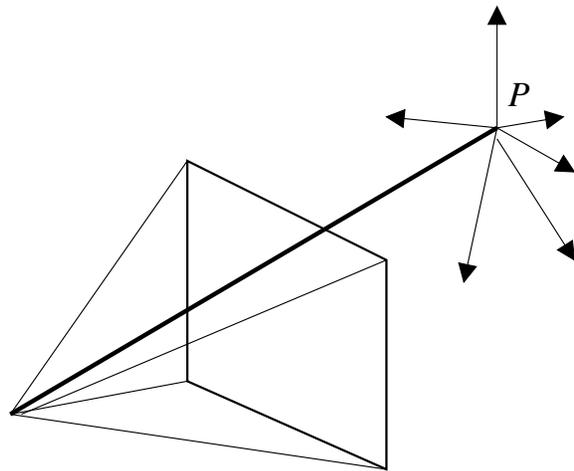
Soit un point 3D \mathbf{P} dans le système de la caméra qui se projète au point image \mathbf{p}' (ou $\mathbf{p} = (x, y)$), donc

$$\mathbf{p}' = \mathbf{JAP} = f \frac{\mathbf{P}}{Z}, \text{ avec } A = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

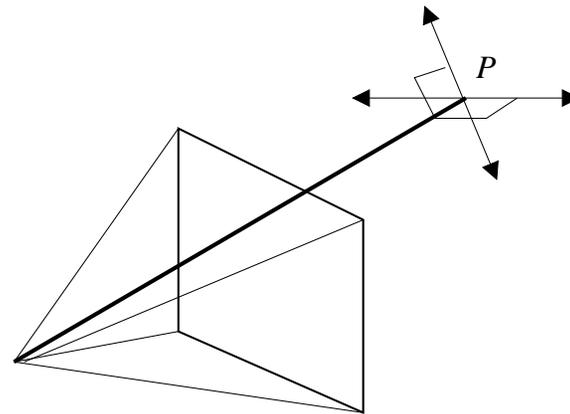
Flux optique d'un objet rigide

Le mouvement relatif \mathbf{V} du point \mathbf{P} par rapport à la caméra est

$$\mathbf{V} = -\mathbf{T} - \omega \times \mathbf{P}$$



Translation T

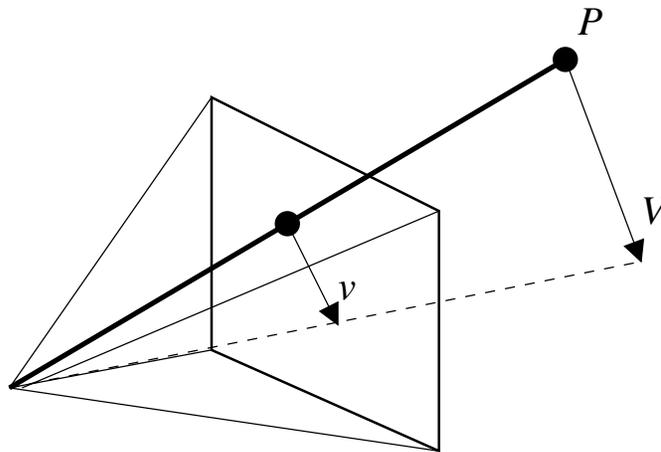


Rotation ω

Flux optique d'un objet rigide

Le mouvement observé \mathbf{v} est le flux optique, et s'obtient en dérivant l'équation de projection $\mathbf{p}' = f \frac{\mathbf{P}}{Z}$ et s'exprime par

$$\mathbf{v} = f \frac{Z\mathbf{V} - V_z\mathbf{P}}{Z^2}$$



ou encore...

$$v_x = \frac{T_z x - T_x f}{Z} - \omega_y f + \omega_z y + \frac{\omega_x x y - \omega_y x^2}{f}$$

$$v_y = \frac{T_z y - T_y f}{Z} - \omega_x f + \omega_z x + \frac{\omega_x y^2 - \omega_y x y}{f}$$

On peut séparer en composante translationnelle et rotationnelle.

$$v_x = v_x^T + v_x^\omega \quad v_y = v_y^T + v_y^\omega$$

Composantes du flux optique

avec

$$v_x^T = \frac{T_z x - T_x f}{Z}$$

$$v_y^T = \frac{T_z y - T_y f}{Z}$$

$$v_x^\omega = -\omega_y f + \omega_z y + \frac{\omega_x x y - \omega_y x^2}{f}$$

$$v_y^\omega = -\omega_x f + \omega_z x + \frac{\omega_x y^2 - \omega_y x y}{f}$$

Propriétés du champs de mouvement

- Le composante rotationnelle v^ω du champs de déplacement ne donne aucune information sur la profondeur dans la scène.
- Le composante translationnelle donne une information sur la profondeur des objets.
- La profondeur Z des objets est liée à la translation T (Si on ne connais si Z ni T , alors on peut récupérer Z et T à un facteur d'échelle près).

Stéréoscopie VS Flux optique

Flux optique

- Mesure *différentielle*.
- Les déplacements spatiaux observés sont des approximations discrètes de dérivées temporelles. Ils doivent donc être petits (typiquement moins de 2 pixels).

Stéréoscopie

- Mesure discrète du déplacement.
- Les déplacements spatiaux observés sont de vrais déplacements. Ils peuvent être de n'importe quelle taille.

Donc la stéréoscopie devient équivalente au flux optique pour des déplacements de caméra très petits.

Flux optique translationnel

Hypothèse

Scène rigide et caméra avec un mouvement translationnel.

On a

$$v_x = \frac{T_z x - T_x f}{Z} \quad v_y = \frac{T_z y - T_y f}{Z}$$

Soit le point $\mathbf{p}_0 = (x_0, y_0)$ tel que

$$\mathbf{p}_0 = (x_0, y_0) = \left(f \frac{T_x}{Z}, f \frac{T_y}{Z} \right)$$

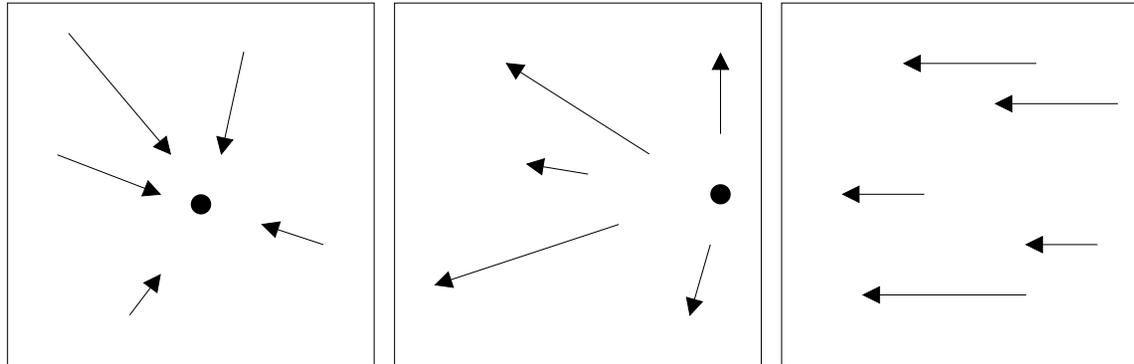
alors on peut substituer et obtenir

$$v_x = (x - x_0) \frac{T_z}{Z} \quad v_y = (y - y_0) \frac{T_z}{Z}$$

donc, un flux optique radial autour du centre (x_0, y_0) .

Flux optique translationnel

Un flux optique radial converge vers le *Point d'expansion*, \mathbf{p}_0 .



- Si $T_z = 0$, le champ est radial dans l'espace projectif et parallèle dans l'espace euclidien.
- Hummm... lien avec les droites épipolaires ?

Flux optique associé à un plan

Hypothèse

Un plan se déplace dans la scène, avec une caméra fixe.

Soit le plan π défini dans le système de la caméra par la relation

$$\mathbf{n}^\top \mathbf{P} = d$$

où \mathbf{n} est la normale du plan et d la position du plan par rapport à l'origine de la caméra.

Le plan se déplace avec une vitesse translationnelle \mathbf{T} et angulaire ω , donc \mathbf{n} et d évoluent dans le temps.

Si on projète le plan sur l'image, on obtient la relation

$$\frac{n_x x + n_y y + n_z f}{f} Z = d$$

Si on isole Z et qu'on ajoute l'équation du flux optique, on obtient

$$v_x = \frac{1}{fd}(a_1x^2 + a_2xy + a_3fx + a_4fy + a_5f^2)$$

$$v_y = \frac{1}{fd}(a_1xy + a_2y^2 + a_6fy + a_7fx + a_8f^2)$$

avec

$$\begin{aligned} a_1 &= -d\omega_y + T_z n_x & a_2 &= d\omega_x + T_z n_y \\ a_3 &= T_z n_z - T_x n_x & a_4 &= d\omega_z - T_x n_y \\ a_5 &= -d\omega_y - T_x n_z & a_6 &= T_z n_z - T_y n_y \\ a_7 &= -d\omega_z - T_y n_x & a_8 &= d\omega_x - T_y n_z \end{aligned}$$

Flux optique associé à un plan

Propriété du champs de mouvement

Le champs de mouvement d'un plan est un polynome quadratique des coordonnées (x,y,f) dans l'image.

- On remarque que le même flux optique peut être crée par deux plans différents avec des mouvements différents.
→ On ne peut pas utiliser un seul plan en mouvement pour calibrer une caméra....

Flux optique et parallaxe

Hypothèse

Supposons que deux points fixes de profondeur différentes de la scène se projettent au même endroit dans l'image et que la caméra se déplace.

Soit les points \mathbf{P} et $\bar{\mathbf{P}}$. Les mouvements sont décrits par

$$\begin{aligned}v_x &= v_x^T + v_x^\omega & v_y &= v_y^T + v_y^\omega \\ \bar{v}_x &= \bar{v}_x^T + \bar{v}_x^\omega & \bar{v}_y &= \bar{v}_y^T + \bar{v}_y^\omega\end{aligned}$$

Si les points se projettent au même endroit, donc $\mathbf{p} = \bar{\mathbf{p}} = (x, y, f)$, alors les vitesses angulaires sont identiques.

En fait, dans les équations

$$v_x^T = \frac{T_z x - T_x f}{Z}$$

$$v_y^T = \frac{T_z y - T_y f}{Z}$$

$$v_x^\omega = -\omega_y f + \omega_z y + \frac{\omega_x x y - \omega_y x^2}{f}$$

$$v_y^\omega = -\omega_x f + \omega_z x + \frac{\omega_x y^2 - \omega_y x y}{f}$$

seul Z est différent, donc seul la composante translationnelle change.

$$\Delta v_x = v_x^T - \bar{v}_x^T \quad \Delta v_y = v_y^T - \bar{v}_y^T$$

On obtient donc

$$\Delta v_x = v_x^T - \bar{v}_x^T = (T_z x - T_x y) \left(\frac{1}{Z} - \frac{1}{\bar{Z}} \right)$$

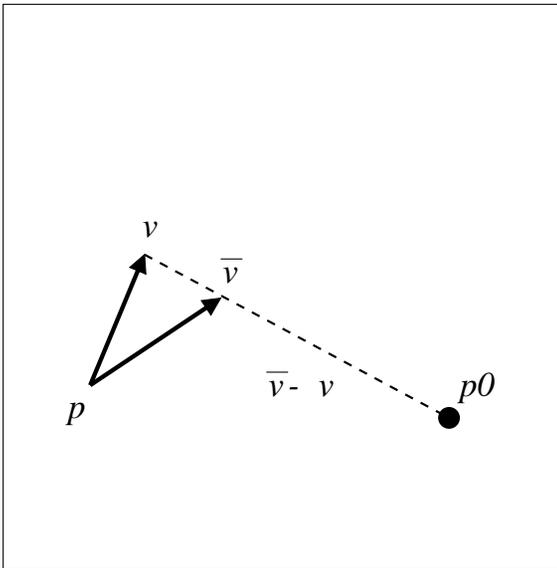
$$\Delta v_y = v_y^T - \bar{v}_y^T = (T_z y - T_y y) \left(\frac{1}{Z} - \frac{1}{\bar{Z}} \right)$$

En divisant, on obtient

$$\frac{\Delta v_x}{\Delta v_y} = \frac{y - y_0}{x - x_0}$$

où \mathbf{p}_0 est le Point d'expansion (*Focus of Expansion*)

$$\mathbf{p}_0 = (x_0, y_0) = \left(f \frac{T_x}{Z}, f \frac{T_y}{Z} \right)$$



Donc, le champs relatif $(\Delta v_x, \Delta v_y)$ pointe vers le point d'expansion \mathbf{p}_0 .

Si on connaît de tels points dans une image, on peut déduire la translation de la caméra.

Reste à estimer le flux optique lui-même....