

IFT 6145
Vision tridimensionnelle
Estimation du mouvement II

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

hiver 2004

Au programme... Le flux optique !

- Chapitre 8 du livre *Trucco & Verri*
 - Hypothèse d'intensité constante (HIC) 8.3.1
 - Problème d'ouverture 8.3.2
 - L'estimation du flux optique 8.4

S.S. Beauchemin et J.L. Barron,
The Computation of Optical Flow

J.L. Barron, D.J. Fleet, S.S. Beauchemin,
Performance of Optical Flow Techniques

- Méthodes différentielles
- Méthodes *domaine fréquences*
- Méthodes *corrélation*
- Approche probabiliste

Estimation du mouvement (Flux Optique)

But :

Déterminer le mouvement de chaque pixel d'une image à l'intérieur d'une séquence d'images

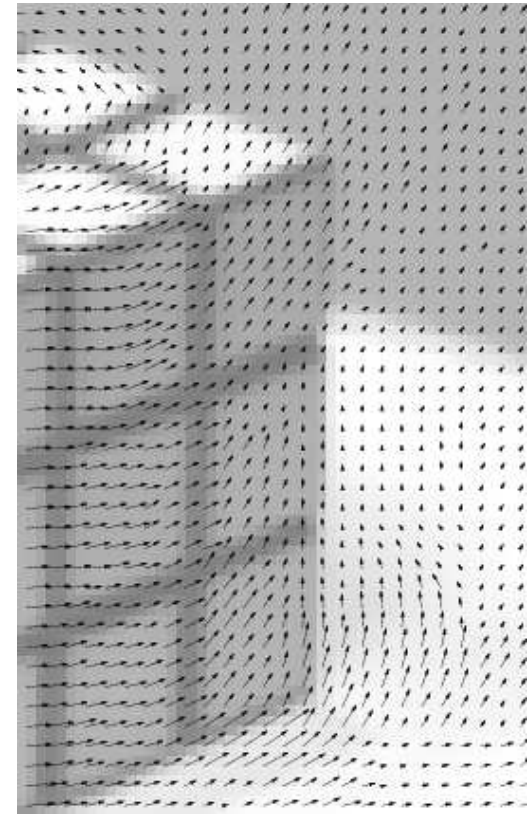
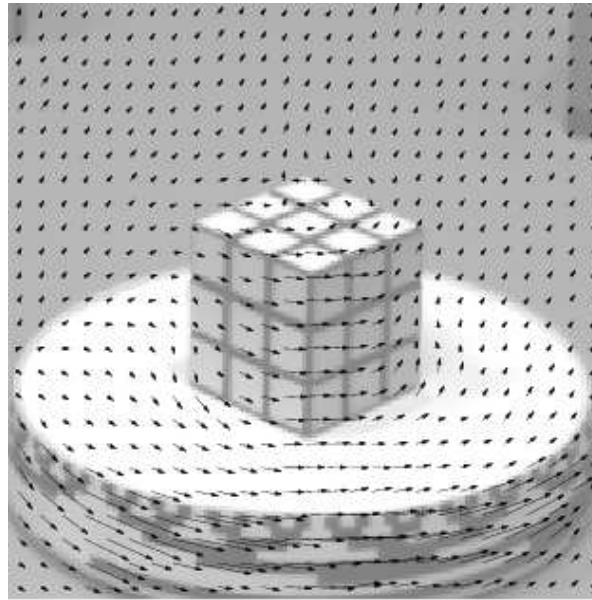
Représenté sous forme d'un champ de mouvement

Hypothèses

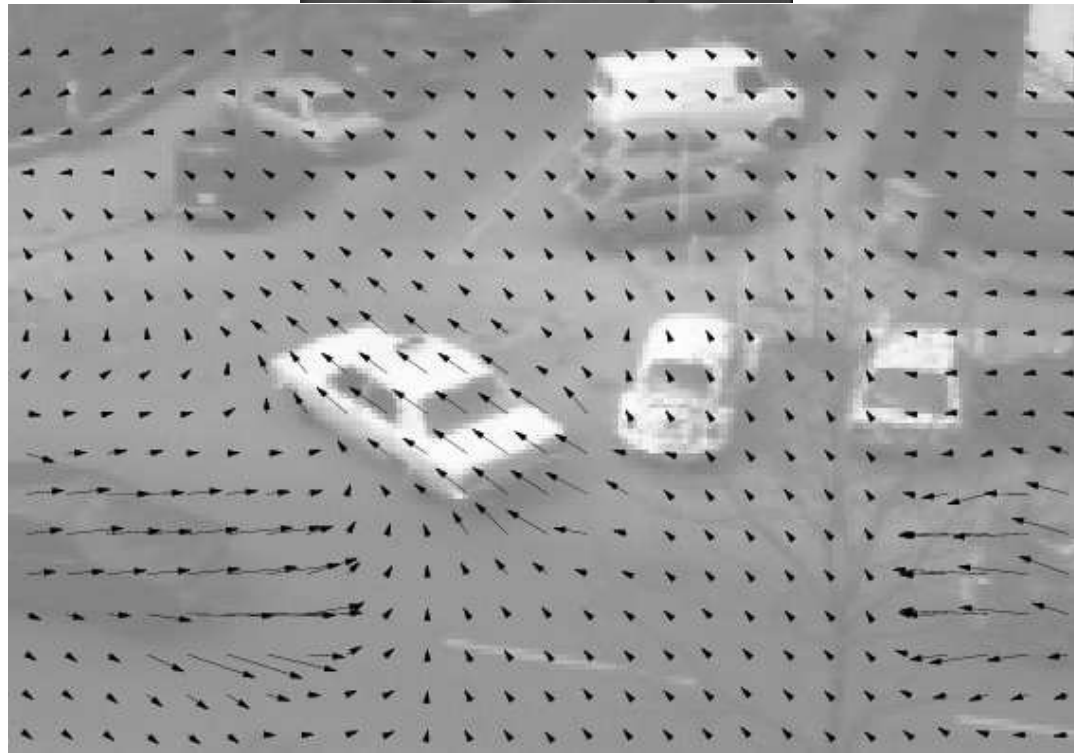
- Mouvement instantané
- Mouvement très petit entre chaque image
- Les changements observés sont uniquement causés par le mouvement

Champ de vitesse (rubic)

Champs induit par le déplacement d'objets dans la scène.



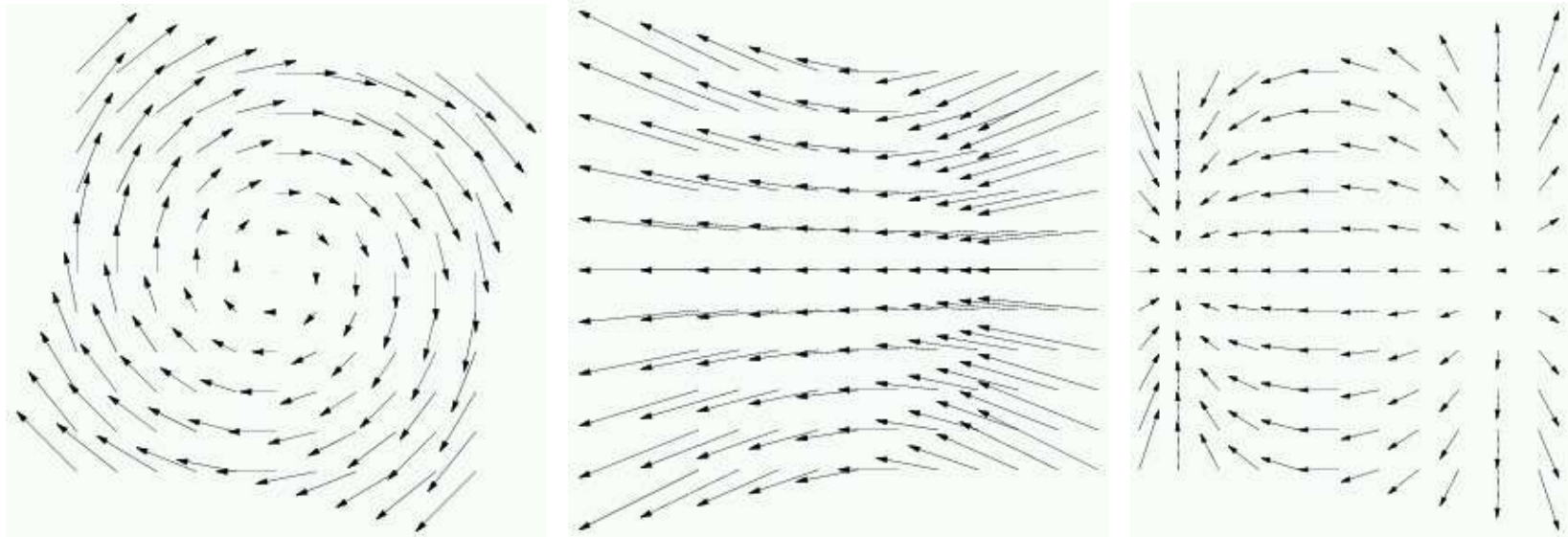
Champs de vitesse (Taxi)



Flux optique et mouvement de caméra

Mouvement de caméra

Estimation de la position et orientation à partir du flux optique



L'approche probabiliste en Vision

Dans le doute, s'abstenir...

- Les probabilités expriment naturellement le manque d'information
- Permet d'augmenter la robustesse en retardant les décisions

Deux raisons d'utiliser les probabilités en Vision

- Problème inverse \Rightarrow ambiguïté

L'approche probabiliste en Vision

Dans le doute, s'abstenir...

- Les probabilités expriment naturellement le manque d'information
- Permet d'augmenter la robustesse en retardant les décisions

Deux raisons d'utiliser les probabilités en Vision

- Problème inverse \Rightarrow ambiguïté

Inverse : Vous avez $a + b = 2$, trouvez a et b

Direct : Vous avez $1 + 1 = c$, trouvez c

L'approche probabiliste en Vision

Dans le doute, s'abstenir...

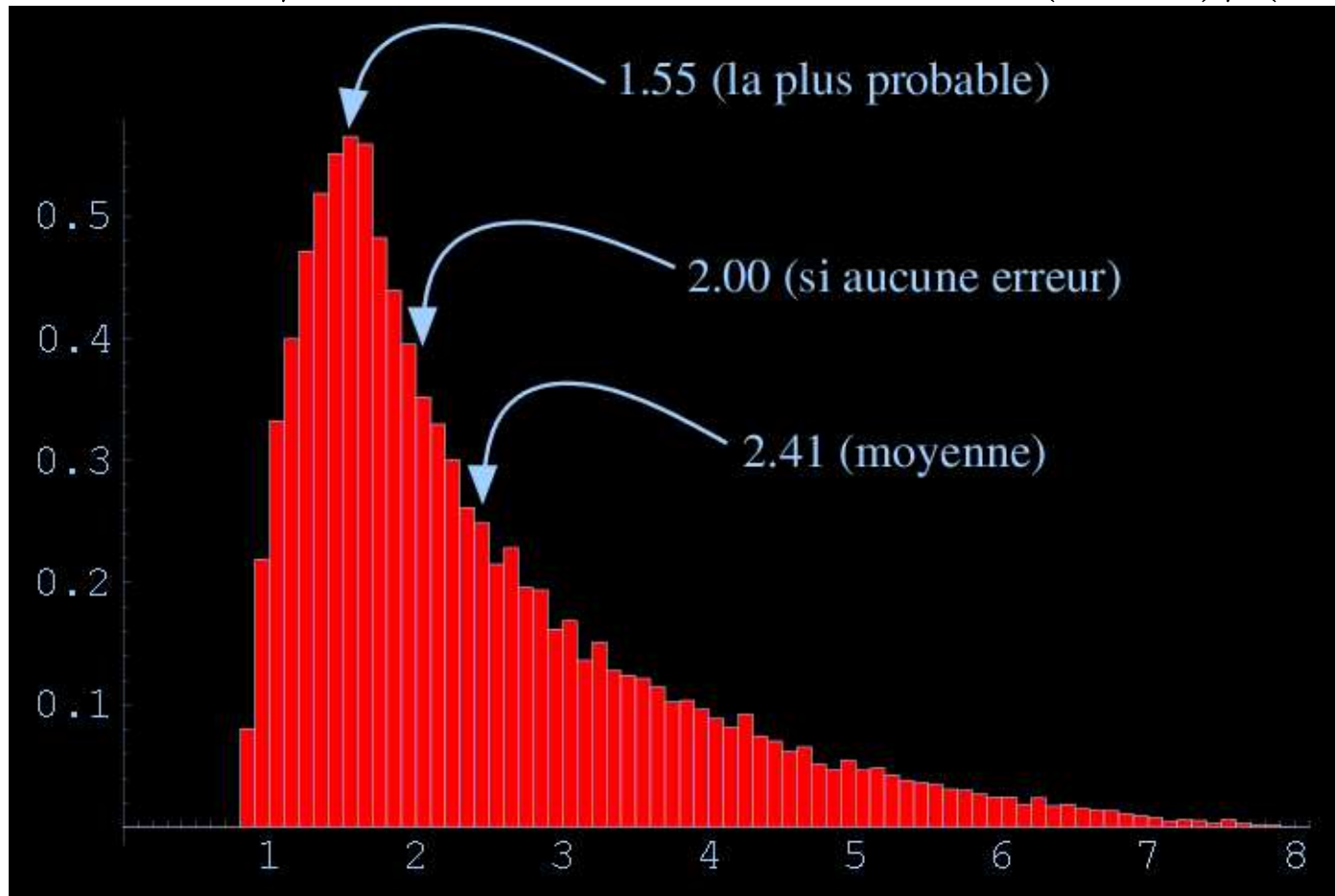
- Les probabilités expriment naturellement le manque d'information
- Permet d'augmenter la robustesse en retardant les décisions

Deux raisons d'utiliser les probabilités en Vision

- Problème inverse \Rightarrow ambiguïté
Inverse : Vous avez $a + b = 2$, trouvez a et b
Direct : Vous avez $1 + 1 = c$, trouvez c
- Le bruit.
Est-ce que $6/3$ est vraiment toujours égal à 2 ?

Approche probabiliste : le problème du bruit

On sait que $6/3 = 2$, mais qu'en est-il de $(6 \pm 2)/(3 \pm 2)$?



Approche probabiliste : le problème du bruit

Quelle valeur choisir pour $(6 \pm 2)/(3 \pm 2)$?

- La valeur “raisonable” est $6/3 = 2$
- La valeur moyenne est 2.41
- La valeur la plus probable est 1.55

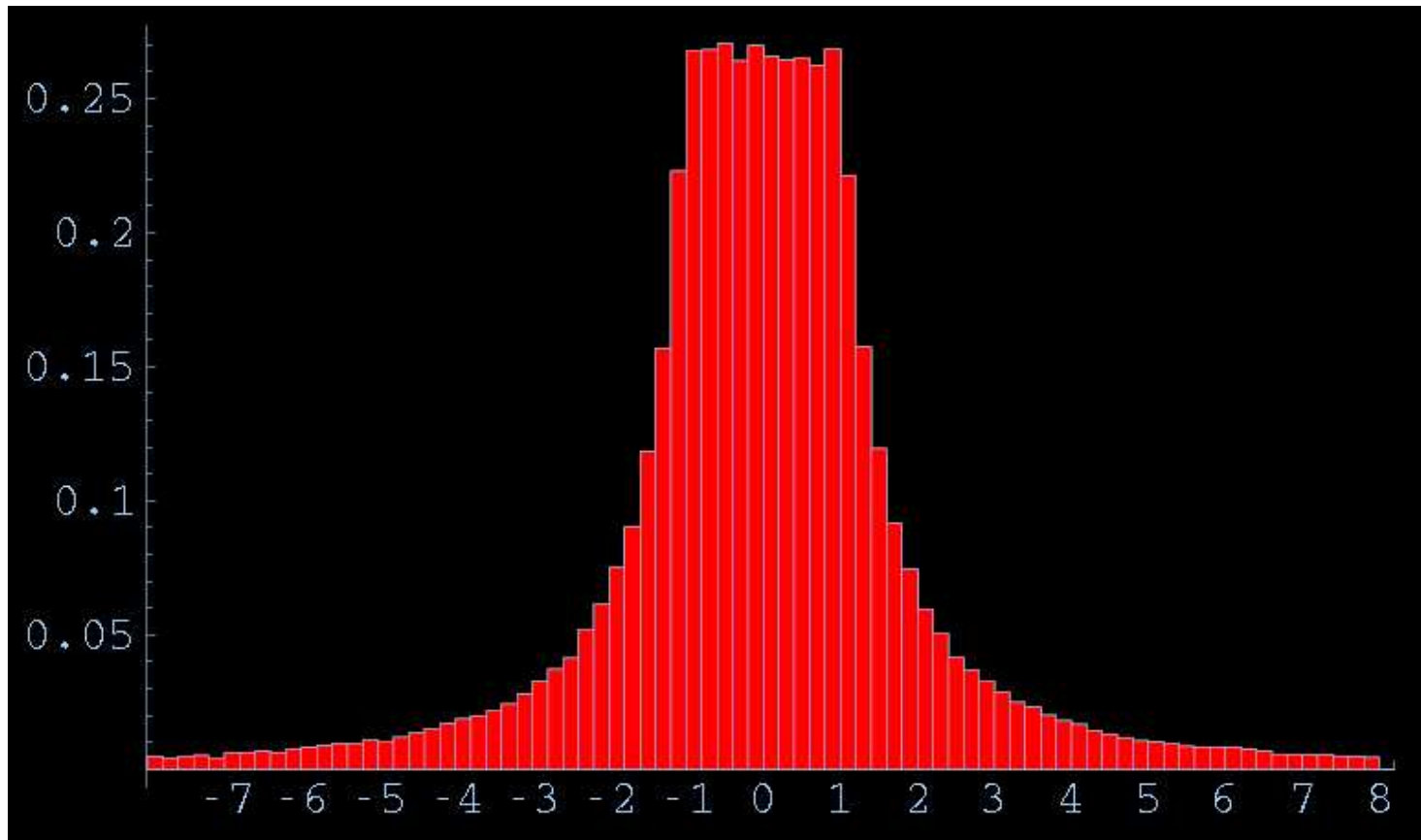
→ **Ça dépend de ce qu’on va faire avec le résultat.**

→ **Mieux vaut conserver la distribution complète.**

Et pour $(0 \pm 2)/(0 \pm 2)$?

Approche probabiliste : le problème du bruit

Distribution pour $(0 \pm 2)/(0 \pm 2)$



Estimation du flux optique

Problème à résoudre

Estimer le champs de vitesse d'une image à partir des seules variations d'intensité des pixels

Trois approches :

- **Approche différentielle**

Utilise les dérivés spatio-temporelles de l'image

- **Approche corrélation**

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

- **Approche domaine fréquence**

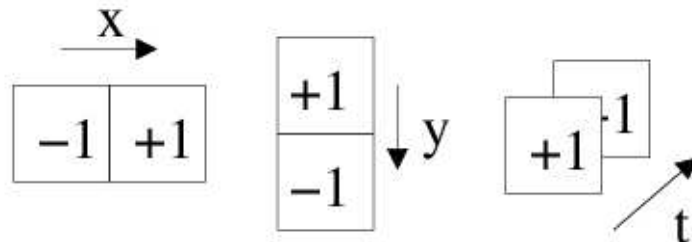
Utilise les changements en fréquence et en phase

Gradient Spatio-temporel

Définition

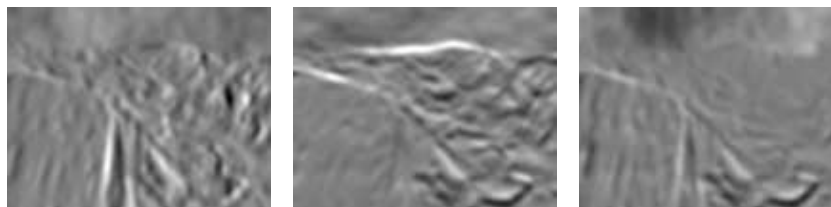
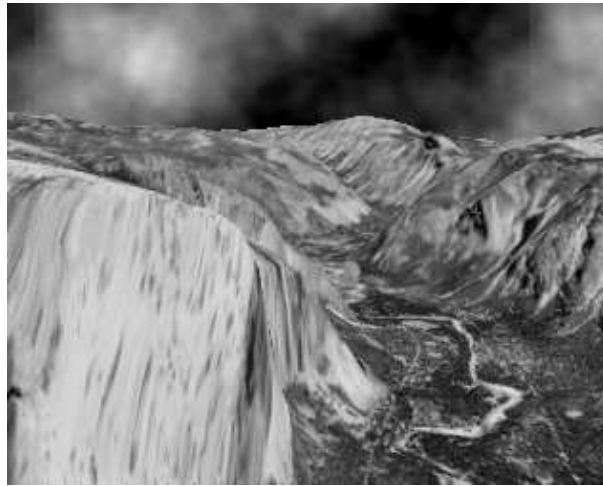
Mesure de la variation d'intensité locale autour d'un pixel

- Mesure “instantanée” pour une seule image
- Gradient spatial : Variations horizontales et verticales (I_x, I_y)
- Gradient temporel : Variations entre deux images successives en (I_t)
- Très utilisé pour détecter les contours



Gradient Spatio-temporel

Séquence et gradients x , y , t .



Les gradients sont en général très bruités...

Hypothèse d'intensité constante

L'intensité observée d'un objet ne change pas lorsqu'il se déplace.

Quand est-ce que c'est respecté en pratique ?

- Objets mats (modèle de réflectance lambertien)
- Source de lumière à l'infini (rayons parallèles)
- Pas de distorsion photométrique dans l'image (centre plus blanc que le bord)
- Pas de distorsion photométriques (caméra parfaite)
- Mouvements faibles (idéalement moins de 1 pixel par image)
- (Pour des détails, voir section 8.3.3)

Hypothèse d'intensité constante (prise 1)

La dérivée temporelle de l'intensité I de l'image doit être nulle, donc

$$\frac{dI}{dt} = 0$$

I est une fonction de (x, y, t) . De plus, x et y sont des fonctions de t . Donc, la dérivée totale de I est

$$\frac{dI(x(t), y(t), t)}{dt} = \frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

Les dérivées partielles $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, et $\frac{\partial I}{\partial t}$ sont les composantes du gradient spatio-temporel et obtenues par convolution de l'image avec des filtres judicieusement choisis...

(hummm... comment choisir ?)

Les dérivées temporelles $\frac{dx}{dt}$ et $\frac{dy}{dt}$ sont les vitesses observées au point (x, y) , donc le flux optique à estimer.

- Gradient spatio-temporel : $\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)$, $I_t = \frac{\partial I}{\partial t}$
- Vitesse du pixel dans l'image : $\mathbf{v} = \left(\frac{dx}{dt}, \frac{dy}{dt} \right)$

On obtient la version simple :

$$\nabla I^\top \mathbf{v} + I_t = 0$$

Si l'intensité change, c'est qu'il y a un déplacement.

Hypothèse d'intensité constante (prise 2)

La notation utilisée est celle de [Beauchemin95].

Si $I(\mathbf{x}, t)$ est l'intensité de l'image au temps t , l'intensité reste constante lors d'un déplacement, donc

$$I(\mathbf{x}, t) = I(\mathbf{x} + \delta\mathbf{x}, t + \delta t)$$

où $\delta\mathbf{x}$ est le déplacement observé dans un interval de temps δt .

L'expansion en série de Taylor donne

$$I(\mathbf{x}, t) = I(\mathbf{x}, t) + \nabla I \cdot \delta\mathbf{x} + \delta t I_t + O^2$$

où $\nabla I = (I_x, I_y)$ et I_t représentent les dérivées partielles du premier ordre de l'image, et O^2 les termes du second degré.

$$I(\mathbf{x}, t) = I(\mathbf{x}, t) + \nabla I \cdot \delta \mathbf{x} + \delta t I_t + O^2$$

En soustrayant $I(\mathbf{x}, t)$ de chaque coté, considérant O^2 négligeable et divisant par δt , on obtient

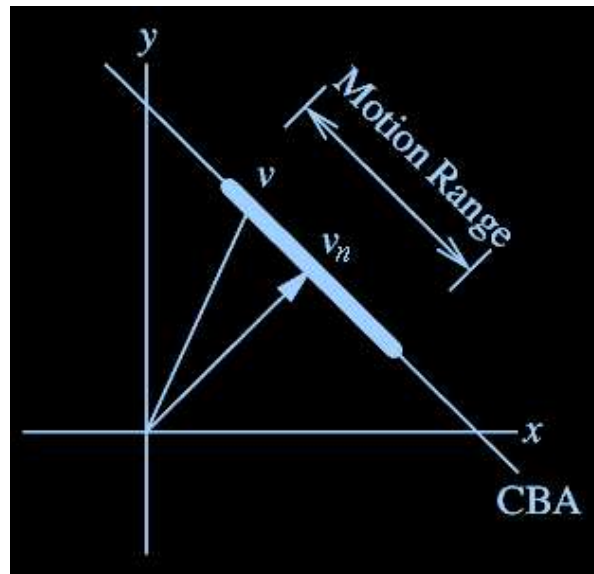
$$\nabla I^\top \cdot \mathbf{v} + I_t = 0$$

où $\mathbf{v} = (u, v)$ est la vitesse observée.

Problème d'ouverture

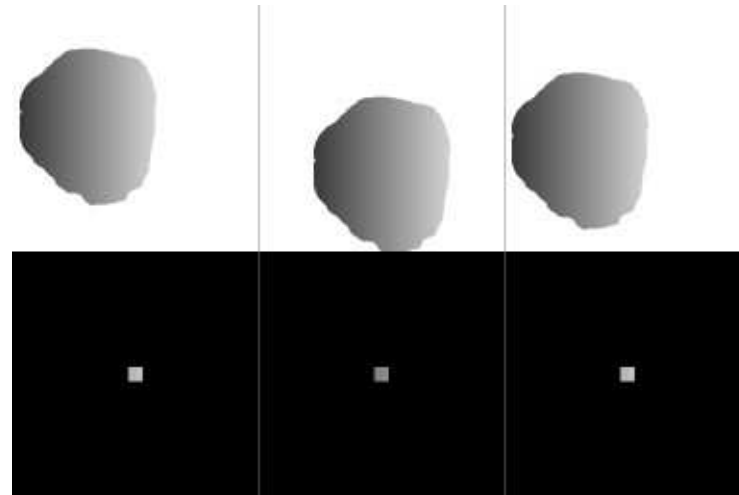
$$\nabla I \cdot \mathbf{v} + I_t = 0$$

ne donne que la composante de la vitesse dans la direction du gradient spatial ∇I . C'est la vitesse normale, \mathbf{v}_n .



Ceci est une manifestation du **problème d'ouverture**...

Problème d'ouverture



Problème d'ouverture

Seule la composante dans le sens du gradient spatial est disponible.

autrement dit...

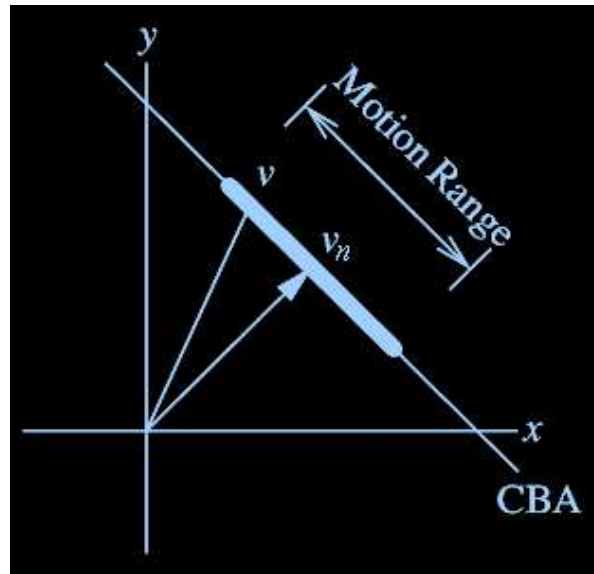
Quand on regarde un objet de trop proche, on ne peut pas estimer exactement sa vitesse.

Flux normal

Le Flux normal est fonction du gradient :

$$\mathbf{v}_n = \frac{-I_t}{\|\nabla I\|^2} \nabla I = \frac{-I_t}{(I_x^2 + I_y^2)} \nabla I$$

Il indique la composante de la vitesse dans la direction du gradient de l'image.



Estimation du flux optique

3 type d'approche

- Différentielles sur l'intensité de l'image
- Fréquences et Phase
- Corrélation

Étapes typiques

1. Préfiltrage spatio-temporel de la séquence
2. Extraction des primitives (ex : dérivées spatio-temporelles)
3. Minimisation d'une fonction d'énergie associée aux primitives.

Estimation du flux optique

Aspects importants à considérer

- Filtrage et différentiation
- Mesure de *confiance*
- Robustesse au type de mouvement

Filtrage et différentiation

Aliasing temporel

- Créé par des déplacements trop grands entre les images.
- Le filtrage temporel est nécessaire.
- Si l'aliasing est trop important, utiliser une approche hiérarchique.

Différentiation numérique

- Pour estimer correctement la dérivée temporelle, on doit utiliser plus de deux images.
- Un simple filtre $[-1, 1]$ n'est pas suffisant.

Mesure de confiance

Confiance basée sur la texture de l'image

- Texture forte → bonne estimation du mouvement
→ grande confiance
- Texture faible → estimation difficile du mouvement
→ confiance moyenne
- Texture nulle → mouvement impossible à estimer
→ aucune confiance.

Utilité de la mesure de confiance

Donne une idée du *bruit* sur l'estimation du mouvement.
Ajoute de l'information pour l'algorithme qui utilisera le mouvement.

Robustesse au type de mouvement

Que faire si le mouvement est grand entre les images ?

- Réduire la taille de l'image ! → réduit les mouvements observés.

Approche hiérarchique ou pyramidale

- On peut réduire successivement jusqu'à ce que le mouvement soit assez petit pour les méthodes différentielles (< 1 pixel par image).
- Le champs de mouvement obtenu à une échelle est appliqué (*warping*) avant de passer à l'échelle suivante.

Observation amusante

On peut réduire le mouvement en réduisant l'image, mais pas augmenter le mouvement en agrandissant l'image. Donc, pas de *pyramide inversée* pour la corrélation...

Estimation du flux optique : Constant-flow

(Algorithme CONSTANT-FLOW de Trucco & Verri)

Hypothèse

- Le flow normal (\mathbf{v}_n) est estimé et considéré *correct*
- Le mouvement est approximativement un champs constant sur une petite région de l'image.

Soit une petite région Q de l'image de $N \times N$ pixels. Pour chaque point \mathbf{p}_i de la région, on peut écrire

$$\nabla I(\mathbf{p}_i)^\top \mathbf{v} + I_t(\mathbf{p}_i) = 0$$

où les gradient $\nabla I(\mathbf{p}_i)$ et $I_t(\mathbf{p}_i)$ sont calculés pour chaque pixel.

Pour chaque point \mathbf{p}_i on a une équation de la forme

$$I_x(\mathbf{p}_i)v_x + I_y(\mathbf{p}_i)v_y = -I_t(\mathbf{p}_i)$$

donc le système linéaire

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{N^2}) & I_y(\mathbf{p}_{N^2}) \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} -I_t(\mathbf{p}_1) \\ -I_t(\mathbf{p}_2) \\ \vdots \\ -I_t(\mathbf{p}_{N^2}) \end{bmatrix}$$

qui se résoud par moindre carré (SVD).

Résumé de l'algorithme “flow constant”

1. Filtrer la séquence d'image avec un filtre spatio-temporel passe-bas gaussien pour réduire le bruit.
2. Filtrer la séquence d'images avec des filtres de dérivée partielles (en x , y et t) pour obtenir les gradients spatio-temporels.
3. Pour chaque pixel, construire et résoudre le système d'équations linéaires pour obtenir v .

Notes

- Le filtrage est spatio-temporel ! ! ! ! Il faut bien comprendre que l'on travaille avec une seule image à trois dimensions (x, y, t) .
- Les étapes 1 et 2 peuvent être combinées.

Horn & Schunck

- Le grand *classique* des méthodes différentielles

Principe

Ajouter un terme lissage qui tend à réduire la dérivé spatiale du champ de mouvement.

$$\int_D \left((\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2 (u_x^2 + u_y^2 + v_x^2 + v_y^2) \right) d\mathbf{x}$$

où D est le domaine d'intérêt, $\mathbf{v} = (u, v)$, et (u_x, u_y, v_x, v_y) sont les dérivés partielles des vitesses.

Comment résoudre ?

Par le Calcul des variations, bien sûr !

En pratique : On procède par itérations successives.

$$u^{k+1} = u'^k - \frac{I_x [I_x u'^k + I_y v'^k + I_t]}{\lambda^2 + I_x^2 + I_y^2}$$
$$v^{k+1} = v'^k - \frac{I_y [I_x u'^k + I_y v'^k + I_t]}{\lambda^2 + I_x^2 + I_y^2}$$

k est le numéro d'itération

(u^0, v^0) est la vitesse initiale définie comme 0

(u'^k, v'^k) est la moyenne des vitesses de l'étape k dans un voisinage

λ le facteur de convergence.

Typiquement...

$\lambda = 0.5$, 100 itérations au moins....

(problème : le nombre d'itération imprévisible)

Horn & Schunck modifié

Pour modéliser explicitement l'expansion/compression d'une image subissant une transformation affine, l'équation

$$\nabla I \cdot \mathbf{v} = -I_t$$

peut être modifié pour donner

$$\nabla I \cdot \mathbf{v} + I(u_x + v_y) = -I_t$$

[On arrête ici pour cette méthode...]

Nagel

Pourquoi se contenter de dérivées premières ?

On permet plus de flexibilité au champs de vitesse en minimisant

$$\int_D \left((\nabla I \cdot \mathbf{v} + I_t)^2 + \frac{\lambda^2}{\|\nabla I\|^2 + 2\delta} \delta (u_x^2 + u_y^2 + v_x^2 + v_y^2) + G_u + G_v \right)$$

avec $G_u = (u_x I_y - u_y I_x)$ et $G_v = (v_x I_y - v_y I_x)$.

La solution itérative est similaire à Horn & Schunk mais contient des dérivés secondes du champs de vitesse et de l'image.

Problème

Les dérivés secondes sont **très** bruitées.

Mukawa

On peut modéliser un modèle d'illumination spéculaire et diffus pour une source de lumière.

La fonction à minimiser est

$$E = \sum_R (u_x^2 + u_y^2 + v_x^2 + v_y^2) + \lambda \sum_R (q + \nabla I \cdot \mathbf{v} + I_t)^2 \\ + \mu \sum_R ((q_x - cI_x)^2 + (q_y - cI_y)^2) + \rho \sum_R (c_x^2 + c_y^2)$$

- 1 Terme de lissage H&S
- 2 Terme d'intensité constante avec q pour composante spéculaire
- 3 Terme reliant l'intensité dans l'image et l'illumination
- 4 Terme assurant le lissage de la fonction d'illumination

Lucas & Kanade

Modèle paramétrique local du type *vitesse constante dans un voisinage*.

Il faut minimiser

$$\sum_{\mathbf{x} \in D} W(x)^2 (\nabla I(\mathbf{x}, t) \cdot \mathbf{v} + I_t(\mathbf{x}, t))^2$$

où $W(\mathbf{x})$ est une fonction de fenêtrage sur le voisinage D .

Ceci correspond à trouver l'intersection "moindre carré" entre les lignes de contraintes HIC du voisinage.

(équivalent au "CONSTANT FLOW" décrit précédemment)

Fréquence et Phase

La transformée de Fourier d'une image en mouvement est

$$\hat{I}(\mathbf{k}, \omega) = \hat{I}_0(\mathbf{k})\delta(\mathbf{v} \cdot \mathbf{k} + \omega)$$

où $\hat{I}_0(\mathbf{k})$ est la transformée de Fourier de l'image $I(\mathbf{x}, 0)$ et \mathbf{x} est la position spatiale. δ est le delta Dirac et (\mathbf{k}, ω) représentent les fréquences spatio-temporelles.

La contrainte d'intensité constante devient dans le domaine fréquence

$$\mathbf{v} \cdot \mathbf{k} + \omega = 0$$

- La vitesse est fonction des fréquences spatiotemporelles.

Orientation spatio-temporelle

Détecter le mouvement = Estimer l'orientation spatiotemporelle

Le filtre de Gabor est très utile. C'est une gaussienne multipliée par un sinus ou un cosinus.

$$G(\mathbf{x}, t) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_t} e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{t^2}{2\sigma_t^2}\right)} \sin(2\pi(\mathbf{x} \cdot \mathbf{k} + \omega t))$$

est un filtre dont la réponse maximale correspond à la fréquence spatio-temporelle (\mathbf{k}, ω) .

Algorithme Fleet & Jepson : Appliquer un filtre de Gabor pour un grand nombre de fréquences spatio-temporelles et utiliser les réponses pour déduire le mouvement.

Modèles paramétriques

On peut remplacer la contrainte de lissage par une contrainte plus rigide... Deux exemples :

– Modèle affine

$$u(x, y) = a_1(x, y) + a_2x + a_3y$$

$$v(x, y) = a_4(x, y) + a_5x + a_6y$$

– Modèle de surfaces localement planaires

$$u(x, y) = a_1(x, y) + a_2x + a_3y + a_7x^2 + a_8xy$$

$$v(x, y) = a_4(x, y) + a_5x + a_6y + a_9xy^2 + a_{10}y^2$$

Flux normal : vers une approche probabiliste

Le Flux normal est fonction du gradient :

$$\mathbf{v}_n = \frac{-I_t}{\|\nabla I\|^2} \nabla I = \frac{-I_t}{(I_x^2 + I_y^2)} \nabla I$$

On sait qu'il y a du bruit sur I , donc il y aura du bruit sur les gradients (I_x, I_y, I_t) ...

Et donc il y aura du bruit sur le flux normal \mathbf{v}_n ...

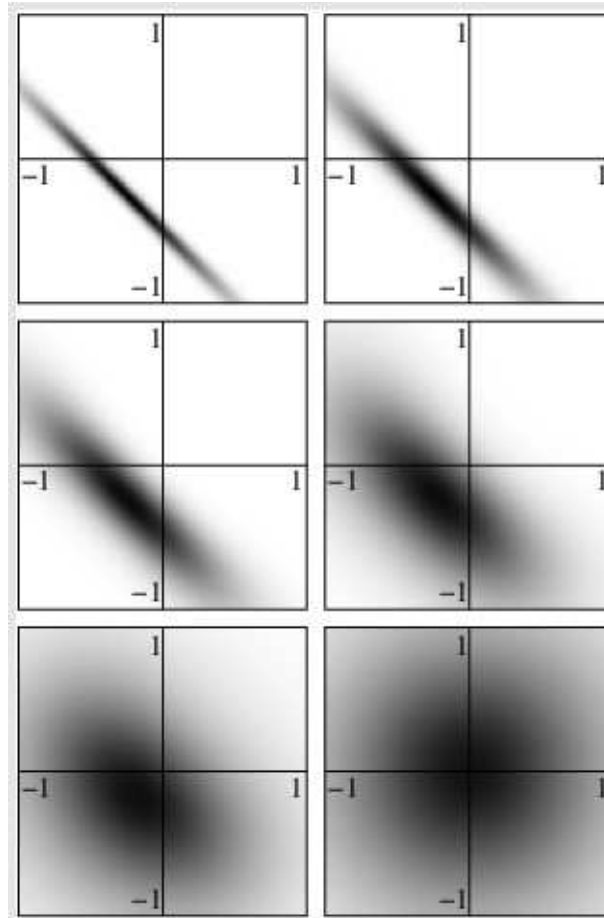
Et donc il y aura du bruit sur le flux optique \mathbf{v} .

Ça rappelle 6/3 ...

Gradient \rightarrow Flux Normal \rightarrow Flux Optique \rightarrow ...

$$p(I_x, I_y, I_t) \rightarrow p(\mathbf{v}_n) \rightarrow p(\mathbf{v}) \rightarrow \dots$$

Distribution de la vitesse v



Gradient spatial $(1, 1, \frac{1}{2})\lambda$ pour $\lambda = [16, 8, 4, 2, 1, 0]$.

Bruit du gradient $\sigma = (1, 1, 1)$.

Résoudre le flux optique

Comment choisir la bonne vitesse dans la distribution, pour chaque pixel ?

⇒ on ajoute une densité jointe qui dit :

Les pixels voisins ont des vitesses similaires

- On ajoute une contrainte de lissage
- On résoud l'orientation de la vitesse en premier
- On résoud la magnitude en second

Et voilà !

Comment on résoud ? Par flot maximum dans les graphes...

Orientation et Magnitude

On exprime $p(\mathbf{v})$ en coordonnées polaires : $p(\theta; m)$.

Orientation θ et magnitude m sont-ils indépendants ?

$$p(v_x; v_y) \neq p(v_x)p(v_y) \quad \text{corr} = 0.4$$

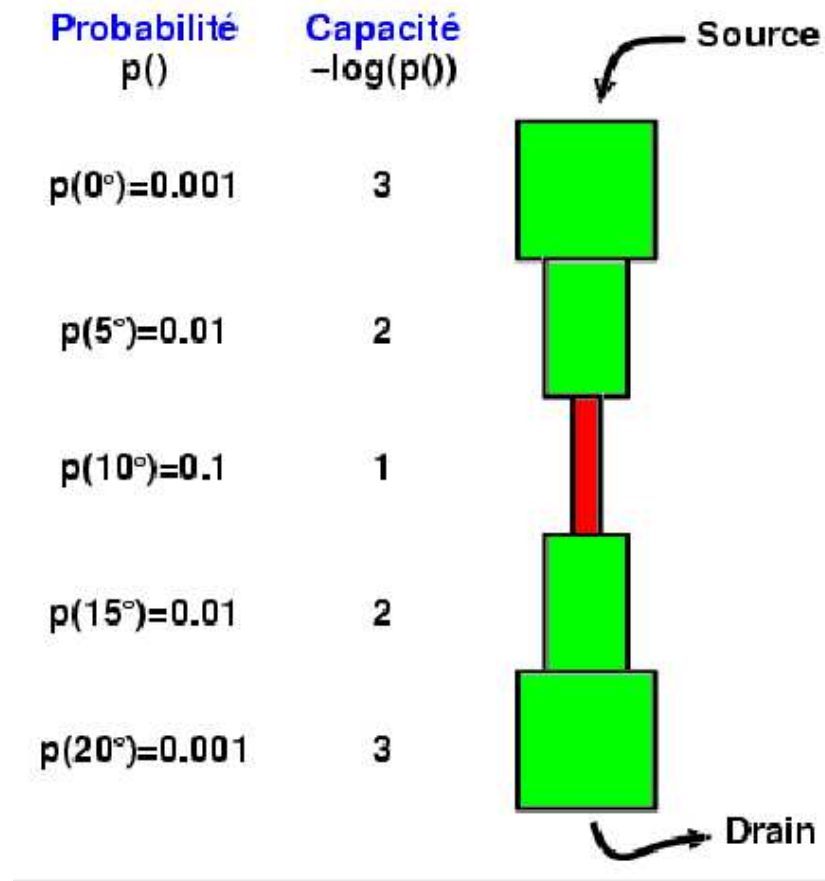
$$p(\theta; m) \approx p(\theta)p(m) \quad \text{corr} = 0.04$$

Oui ! (Presque !)

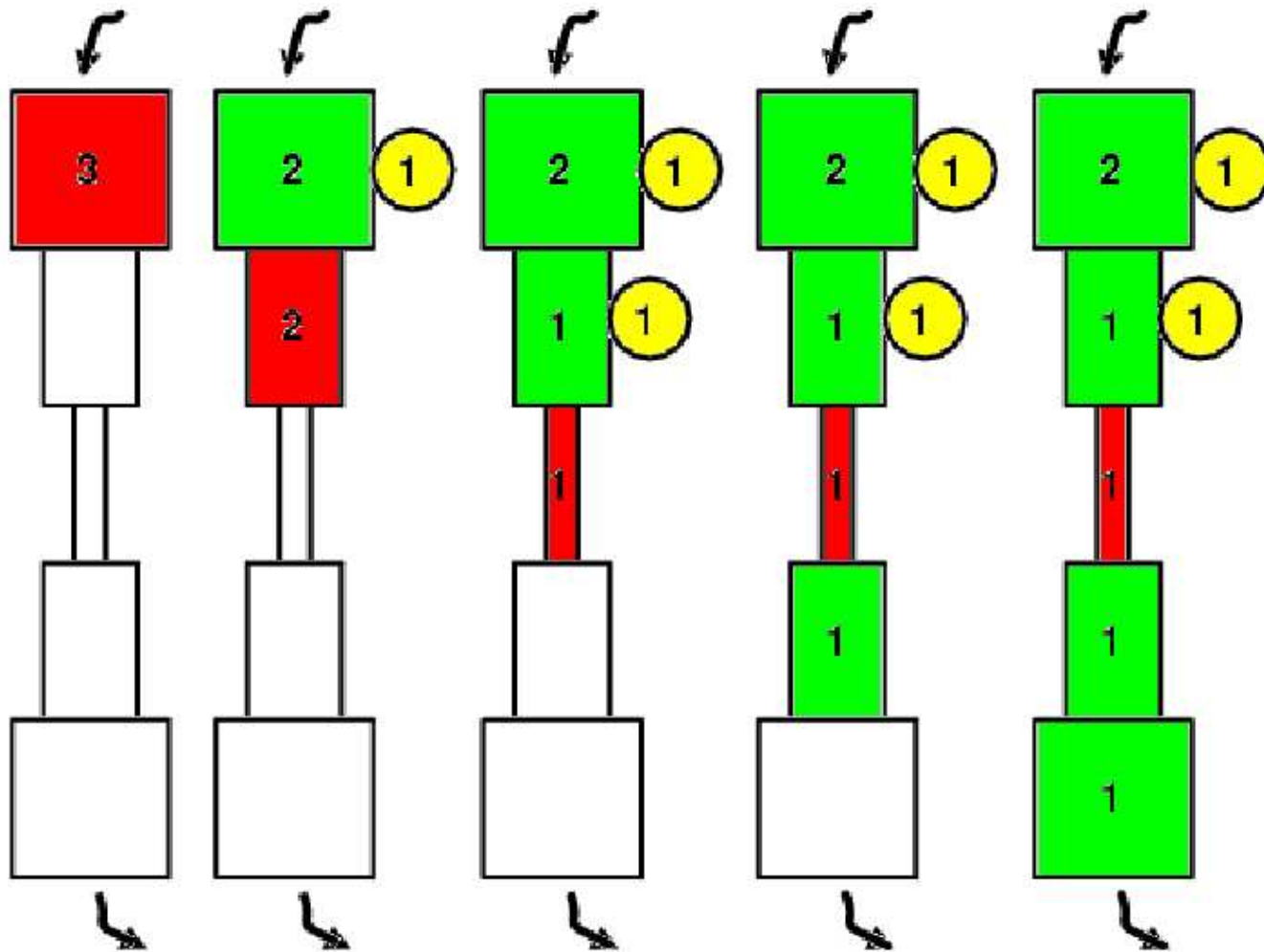
Probabilité et Plomberie

Champs aléatoire de Markov...

Théorème de Hammersley-Clifford...

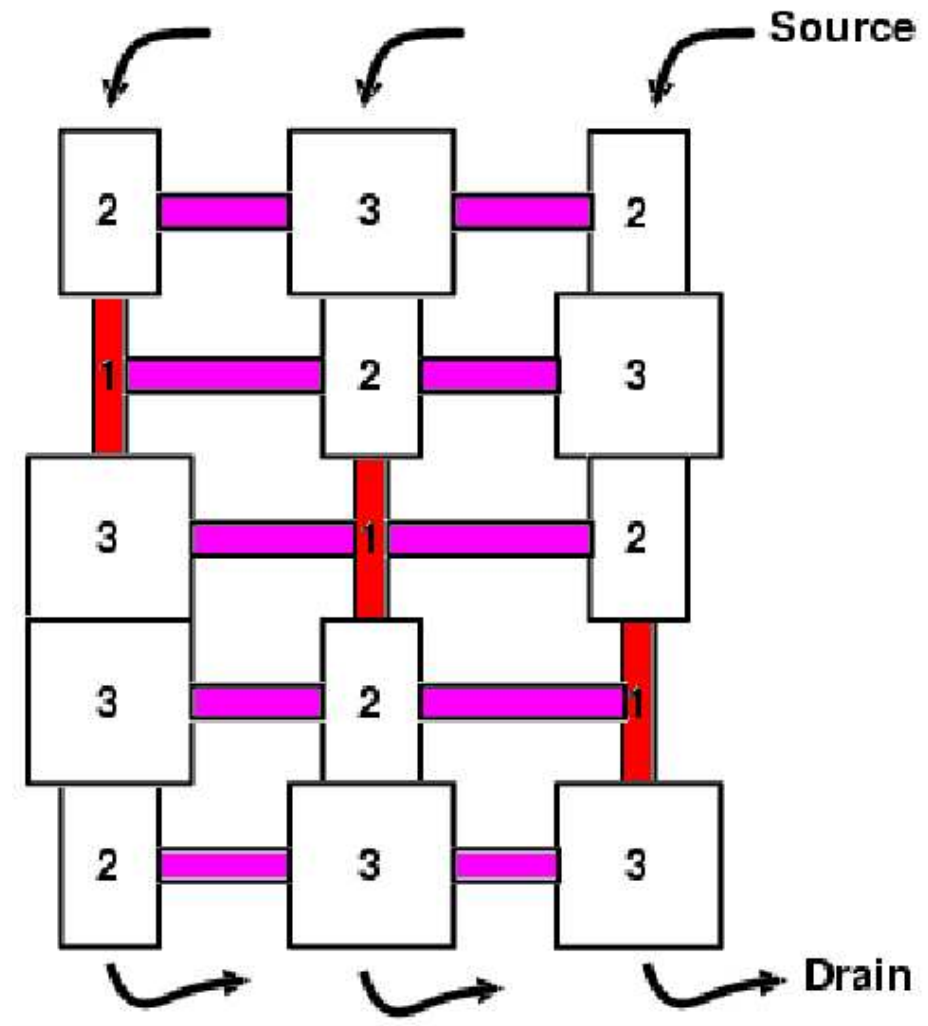


Preflow-push lift-to-front

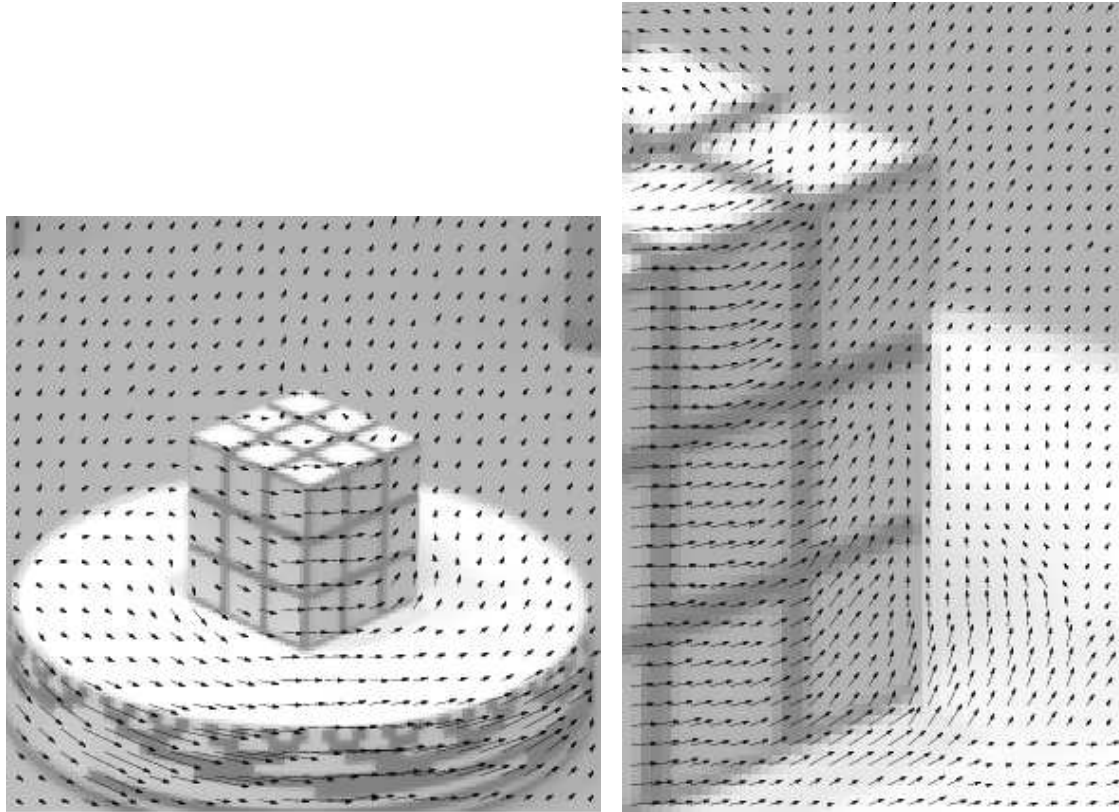


Et le lissage ?

Il suffit d'ajouter quelques tuyaux....

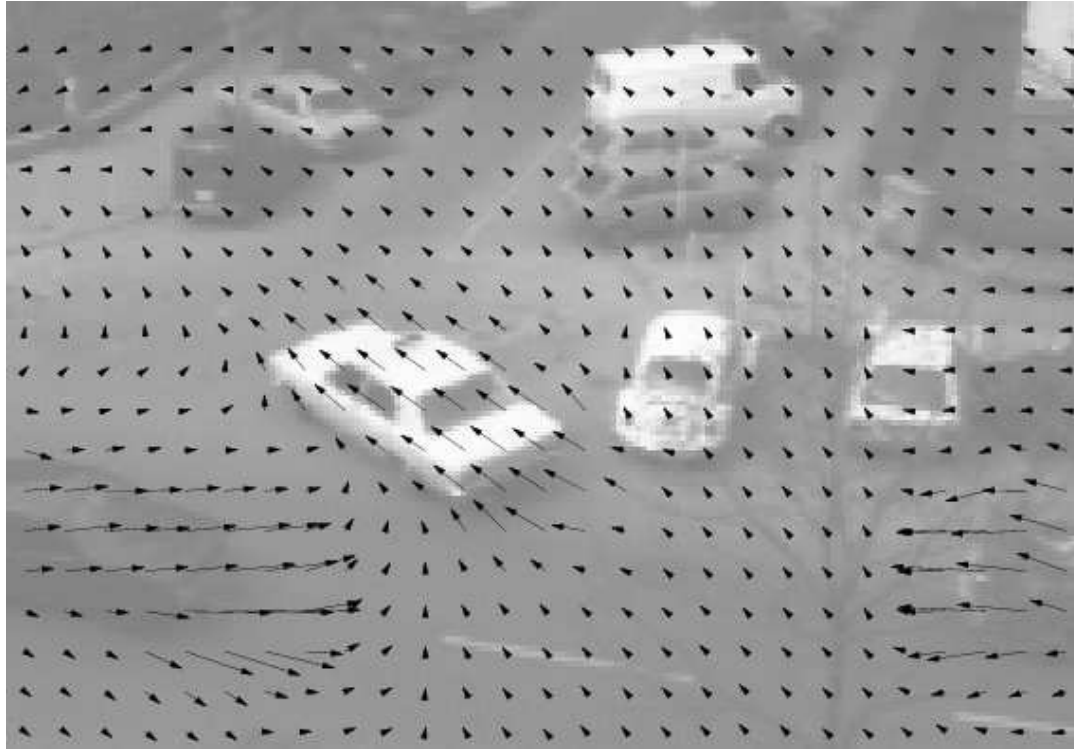


Résultats : rubic



Les zones sans texture sont correctement évaluées.

Résultats : taxi



Les discontinuités du flot sont bien préservées.