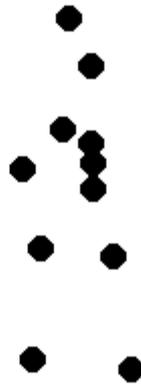
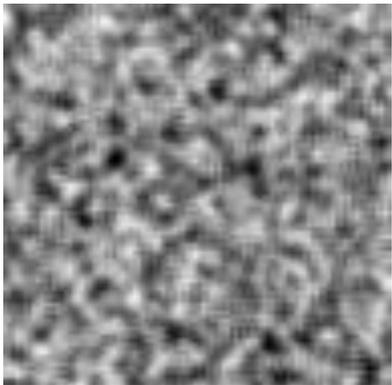


Analyse du mouvement

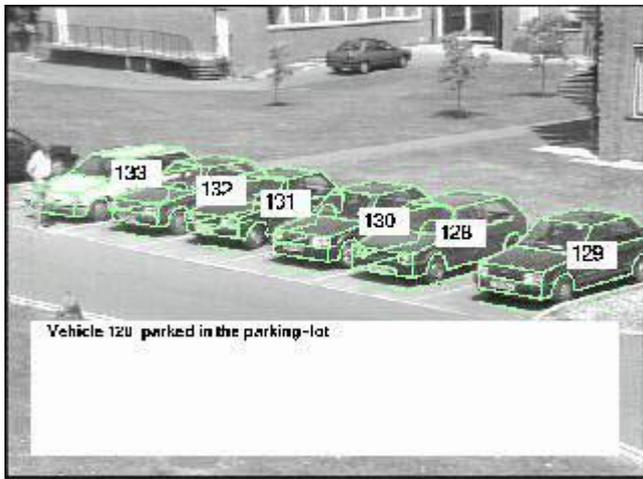
Estimation du flot optique

Antoine MANZANERA – ENSTA / U2IS



Applications et enjeux (1)

Surveillance de zone



(University of Surrey)

Compression et Indexation



Applications et enjeux (2)

Navigation robotique



(Ecole des Mines / CMM)

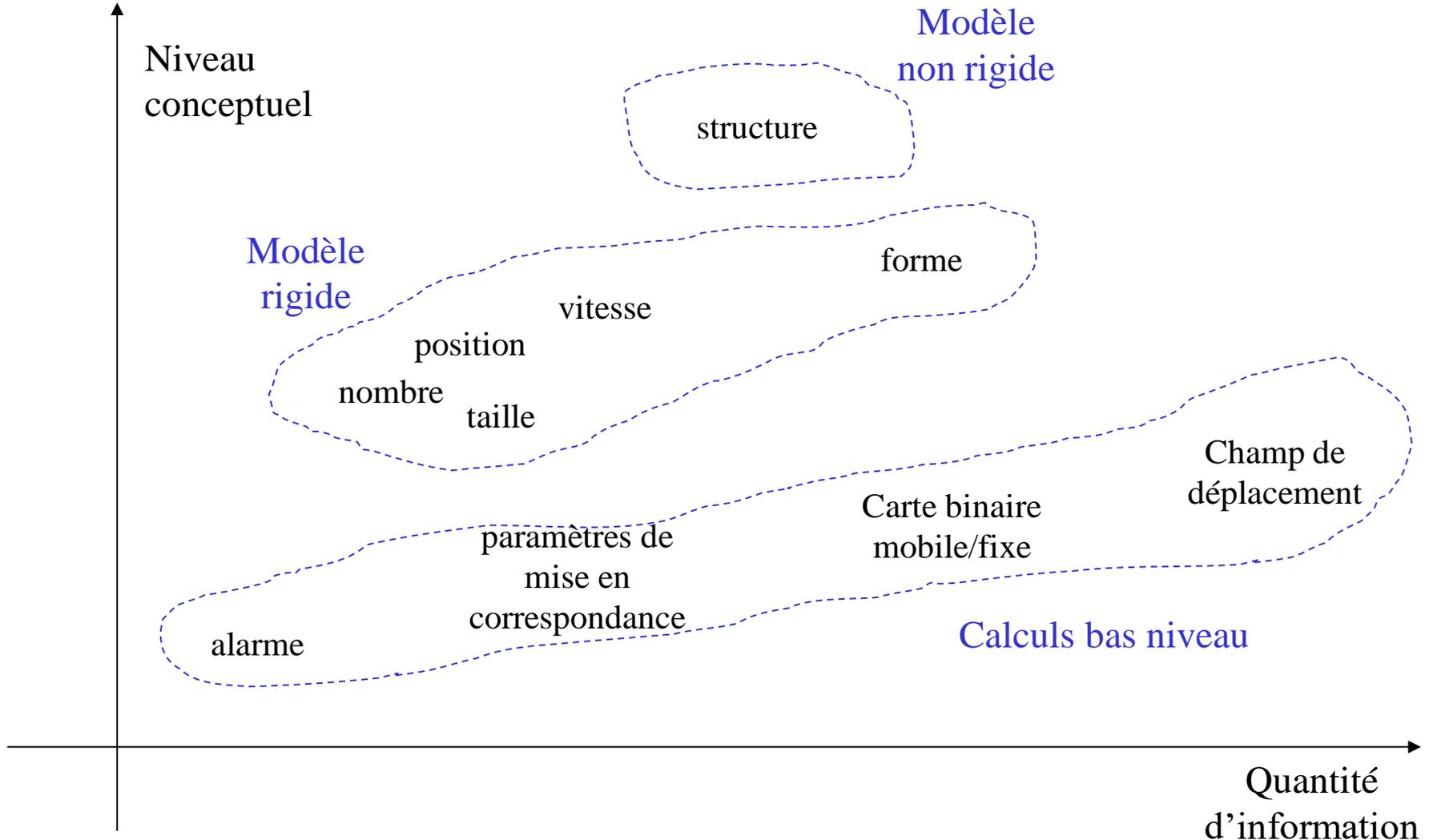
Poursuite automatique



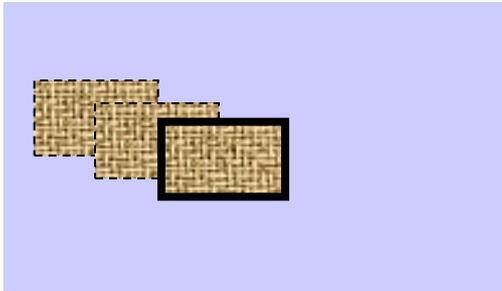
Mouvement fluide



Quelle information extraire ?



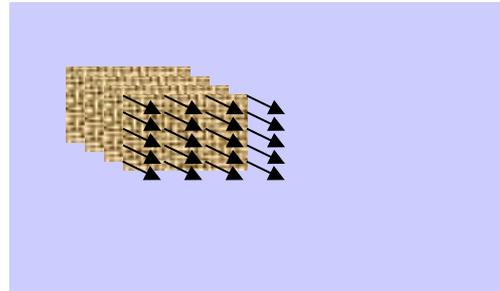
Détection vs Estimation vs Poursuite



DETECTION

Objectif : identifier dans chaque image les pixels appartenant à des objets mobiles

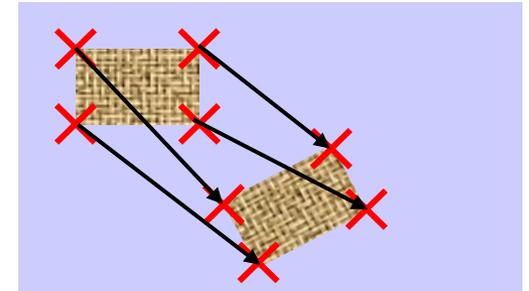
- Certaine continuité temporelle
- Mouvement de la caméra nul ou très simple.



ESTIMATION

Objectif : calculer le mouvement apparent (vitesse instantanée) de chaque pixel

- Continuité temporelle
- Plutôt « traiter après »



POURSUITE

Objectif : apparier certaines structures spatiales pour chaque couple d'images.

- Discontinuité temporelle
- Plutôt « traiter avant »

MOUVEMENT : Estimation

PLAN DU CHAPITRE

- **Champ de vitesse apparent et applications**
- **Techniques par appariement**
- **Techniques différentielles 1 : Lucas et Kanade**
- **Techniques différentielles 2 : Horn et Schunck**
- **Techniques multi-résolutions**
- **Techniques par apprentissage**

Calcul du mouvement apparent

(1) Le calcul d'un mouvement apparent *global* (mise en correspondance) entre deux images correspond à l'estimation des paramètres d'une transformation affectant *tous* les points de l'image : translation, rotation, homothétie, affinité,...

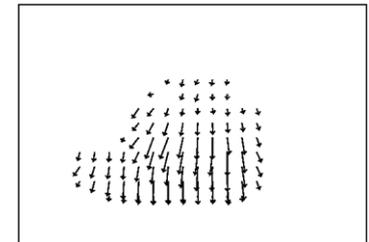
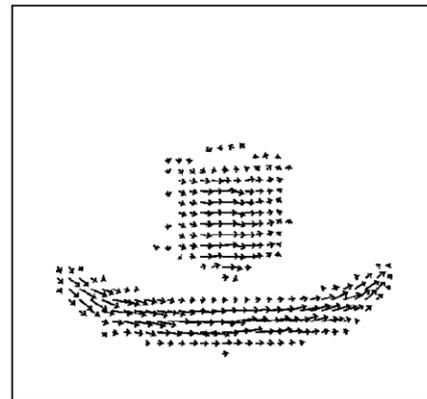
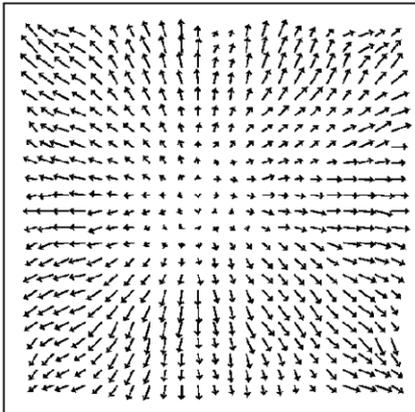
(2) Le calcul du mouvement apparent *local* consiste à associer à chaque pixel (x,y,t) de I un vecteur (v_x^t, v_y^t) représentant la *vitesse apparente* du pixel (x,y) à l'instant t .

—————→ Calcul du *flot optique* (= Champ de mouvement apparent)

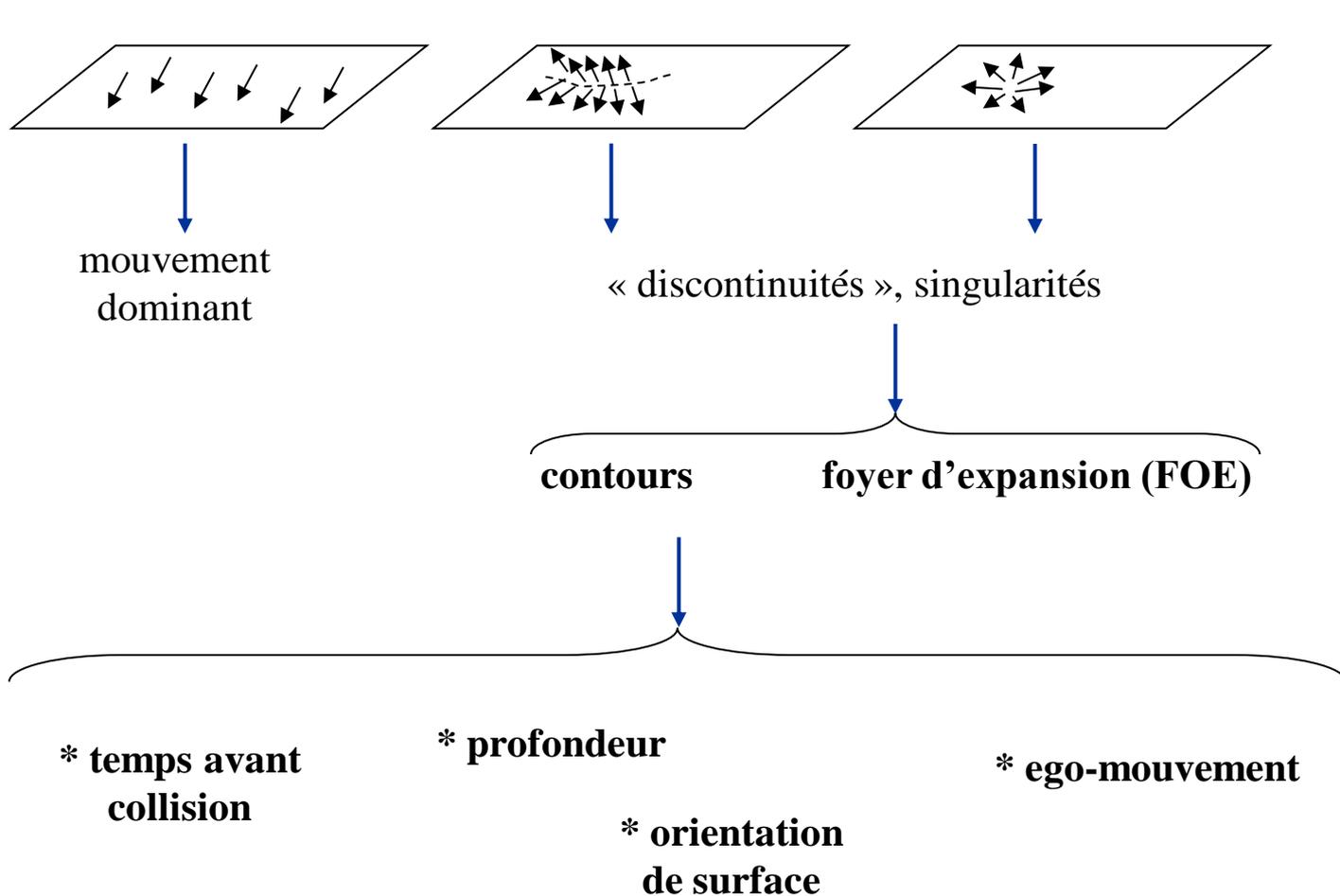
Idéalement : le vecteur (v_x^t, v_y^t) représente la projection sur le plan image du vecteur vitesse (V_x^t, V_y^t, V_z^t) des objets de la scène par rapport au repère image (O,x,y,z) (grandeur objective).

On le calcule à partir des variations temporelles de la fonction $I(x,y,t)$.

Exemples de flot optique

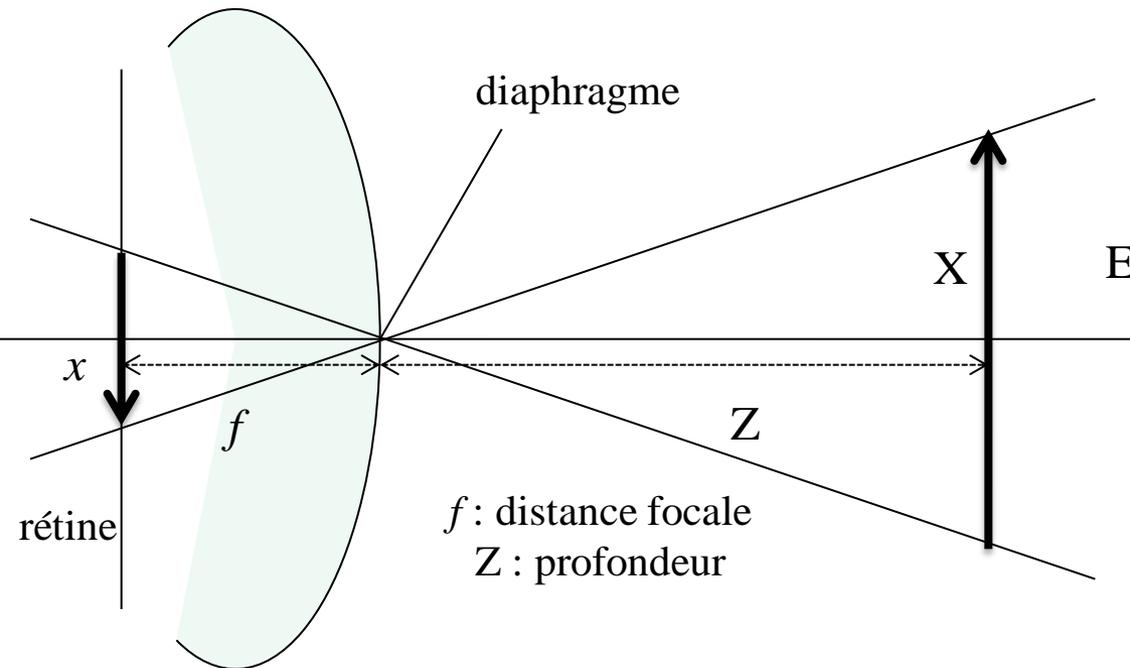


Quelles informations peut fournir le flot optique ?



niveau
conceptuel

Flot optique théorique



(O, X, Y, Z) coordonnées réelles 3d

(O, x, y) coordonnées rétine 2d

Equations perspectives (modèle sténopé) :

$$\begin{cases} x = \frac{fX}{Z} \\ y = \frac{fY}{Z} \end{cases}$$

La dérivée temporelle des équations perspectives fournit celles du flot optique théorique :

$$\begin{aligned} \dot{x} &= f \left(\frac{\dot{X}}{Z} - \frac{X\dot{Z}}{Z^2} \right) \\ \dot{y} &= f \left(\frac{\dot{Y}}{Z} - \frac{Y\dot{Z}}{Z^2} \right) \end{aligned}$$

avec :

$$\dot{x} = \frac{dx}{dt}$$

Flot optique et parallaxe

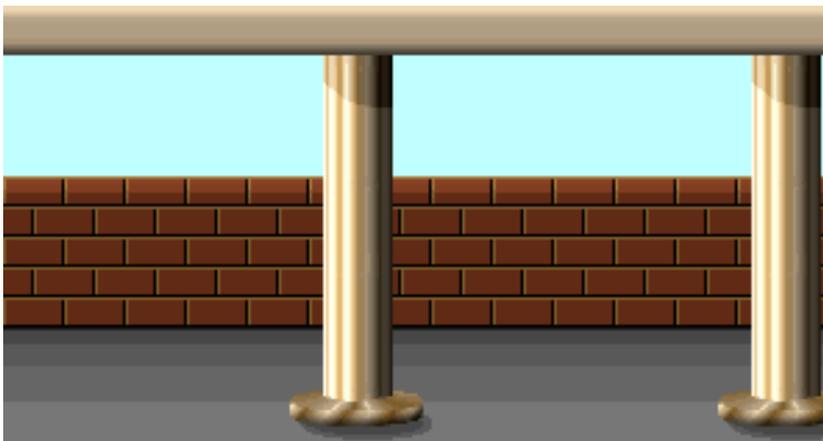
Dans le cas d'une translation pure selon l'axe OX (travelling horizontal), on a :

$$Z \dot{=} 0, \text{ et donc } \dot{x} = \frac{f\dot{X}}{Z}$$

soit :

$$Z = \frac{f\dot{X}}{\dot{x}}$$

La profondeur est inversement proportionnelle au module de la vitesse apparente



[© nvnews.net]

Translation le long de l'axe optique

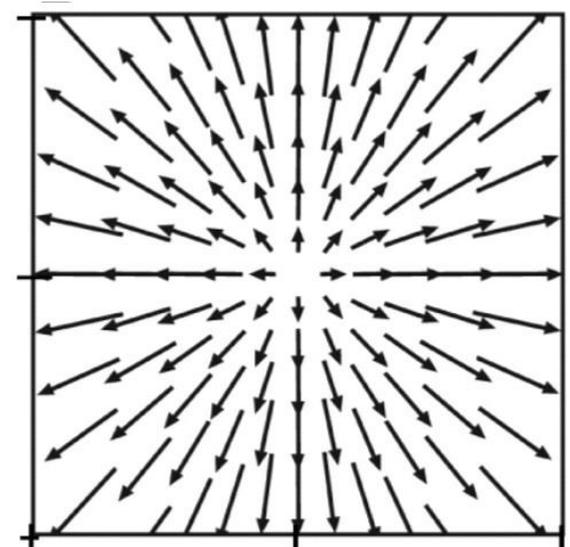
(O,X,Y,Z) coordonnées réelles 3d ; (O,x,y) coordonnées rétine 2d

$$\text{Equations perspective (modèle sténopé) : } \left\{ \begin{array}{l} x = \frac{f X}{Z} \\ y = \frac{f Y}{Z} \end{array} \right. \quad \text{Dérivée temporelle (flux optique) : } \left\{ \begin{array}{l} \dot{x} = f \left(\frac{\dot{X}}{Z} - \frac{X \dot{Z}}{Z^2} \right) \\ \dot{y} = f \left(\frac{\dot{Y}}{Z} - \frac{X \dot{Y}}{Z^2} \right) \end{array} \right.$$

Cas d'un mouvement de translation pure le long de l'axe optique (OZ) :

$$\left\{ \begin{array}{l} \dot{x} = -f \frac{X \dot{Z}}{Z^2} = -\frac{x \dot{Z}}{Z} \\ \dot{y} = -f \frac{X \dot{Y}}{Z^2} = -\frac{y \dot{Z}}{Z} \end{array} \right.$$

Le champ de déplacement forme un zoom sur l'image avec foyer d'expansion (FoE) au niveau du centre optique O.



Généralisation du foyer d'expansion (FOE)

Lors d'une translation de la caméra dans une scène statique, les directions de vitesse des points projetés sur le plan image converge vers un point du plan projectif appelé FOE.

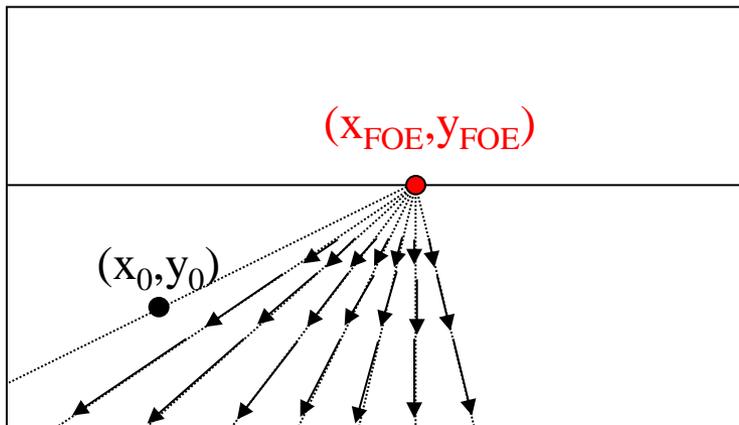
Dans la suite, on suppose que la caméra se déplace selon la translation $T = (-X', -Y', -Z')$. Pour simplifier les notations, on suppose de plus que $f_x = f_y = f$ et $(x_c, y_c) = (0, 0)$.

Soit (X_0, Y_0, Z_0) un point de la scène. Après un temps t , il est projeté sur l'image au point (x_t, y_t) , avec :

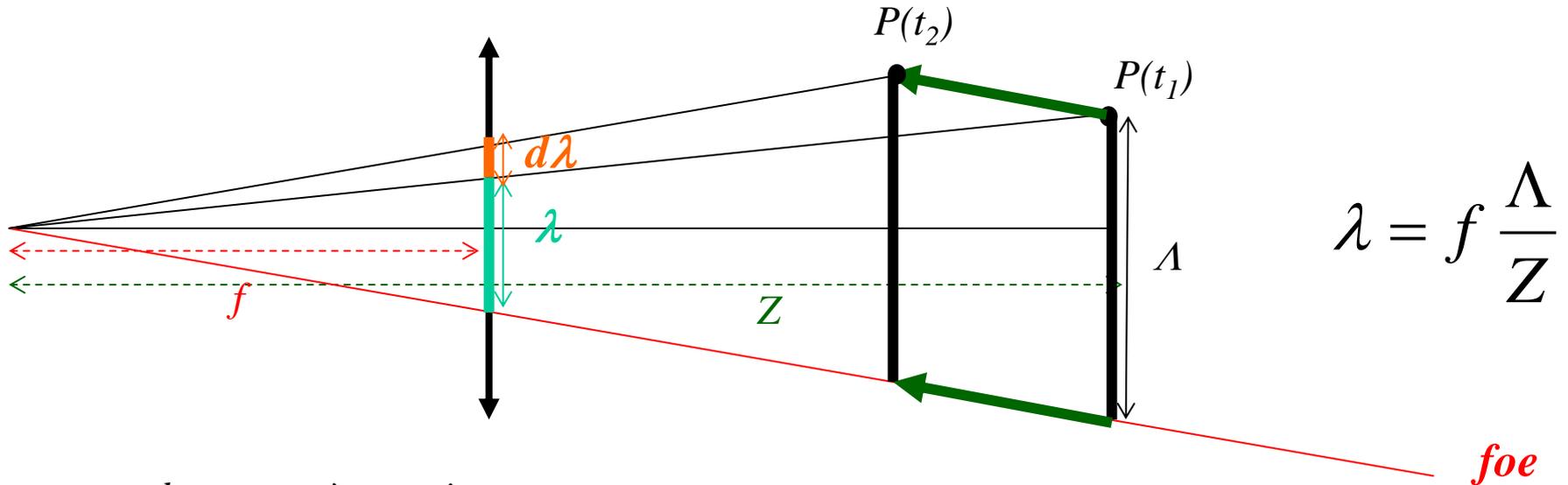
$$(x_t, y_t) = \left(f \frac{X_0 + tX'}{Z_0 + tZ'}, f \frac{Y_0 + tY'}{Z_0 + tZ'} \right)$$

$$(x_{FOE}, y_{FOE}) = \lim_{t \rightarrow -\infty} (x_t, y_t)$$

$$= \left(f \frac{X'}{Z'}, f \frac{Y'}{Z'} \right)$$



Temps avant collision



$$\lambda = f \frac{\Lambda}{Z}$$

mouvement dans une scène statique :

$$\lambda' = -\frac{f\Lambda}{Z^2} Z'$$

et donc

$$\frac{\lambda}{\lambda'} = -\frac{Z}{Z'}$$

temps avant collision

Temps avant collision (suite)

Chez la plupart des espèces animales, y compris chez les humains, on observe un réflexe de retrait face à la projection 2d d'un objet en expansion.

Certains oiseaux (ex : pigeons) ont des neurones dont la réponse est fortement corrélée au temps avant collision.



L'arrivée d'un train en gare de La Ciotat – L. et A. Lumière (1895)

Calcul du flot optique : limites et contraintes

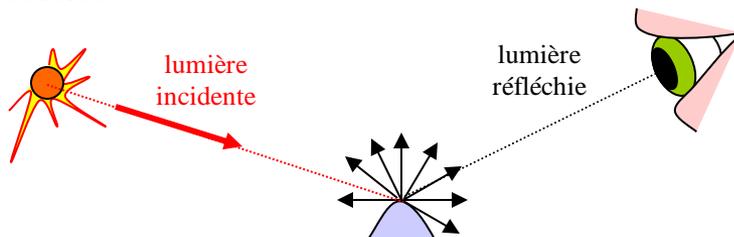


(1) On suppose :

MOUVEMENT \Leftrightarrow VARIATION D'INTENSITE

(\Leftarrow) On suppose que l'observation d'une variation d'intensité dans l'image traduit nécessairement l'existence d'un mouvement dans la scène. Cela correspond à une hypothèse *d'éclairement constant* sur la scène.

(\Rightarrow) C'est la variation de l'intensité dans l'image qui doit permettre de retrouver le mouvement apparent de l'objet dans la scène. On suppose donc que la lumière réfléchie par un point de la scène reste constante indépendamment du mouvement relatif scène / caméra. L'hypothèse sous-jacente est que les objets sont à surface *lambertienne* :



L'intensité de la lumière réfléchie est la même dans toutes les directions.

Calcul du flot optique : limites et contraintes



Barber pole



(2) PROBLEME DE L'OUVERTURE

On n'accède au mouvement apparent d'un point que grâce à un calcul effectué *dans un voisinage borné* de ce point.

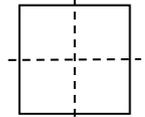
On ne peut donc calculer que la composante du mouvement *dans la direction du gradient* (i.e. *perpendiculaire au contour*).



Technique par appariement

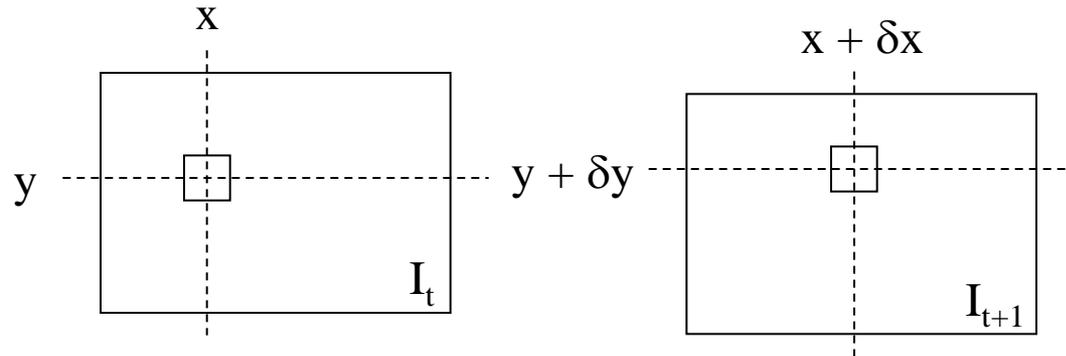
Mesure d'appariement :

On considère $B \subset \mathbf{Z}^2$ voisinage de l'origine :



$$A_{(x,y)}^t(\delta x, \delta y) = \sum_{(b_1, b_2) \in B} (I(x + b_1, y + b_2, t) - I(x + \delta x + b_1, y + \delta y + b_2, t + 1))^2$$

C'est la somme des différences au carré (SSD) de 2 blocs :



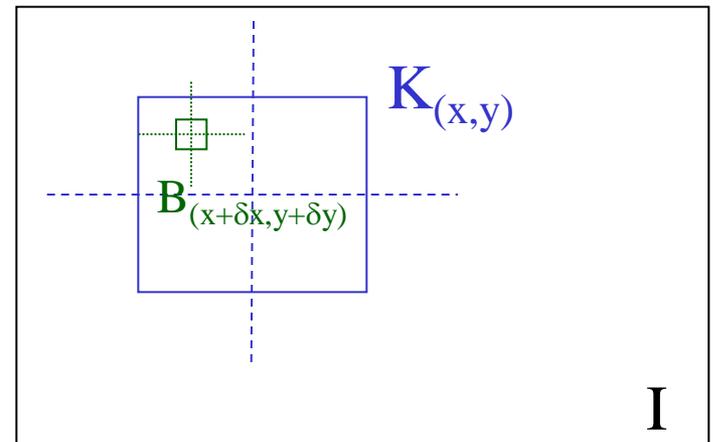
typiquement :
B de taille
9×9, 15×15...

Solution du flot optique :

$$(v_x^t, v_y^t) = \arg \min_{(\delta x, \delta y) \in K} A_{(x,y)}^t(\delta x, \delta y)$$

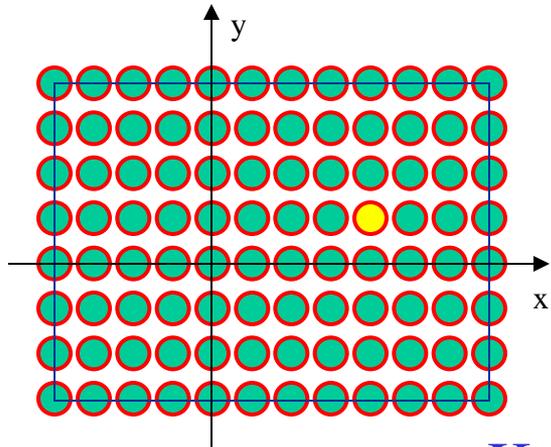
$$K \subsetneq I$$

pour limiter le temps de calcul
+ stratégies d'optimisation...



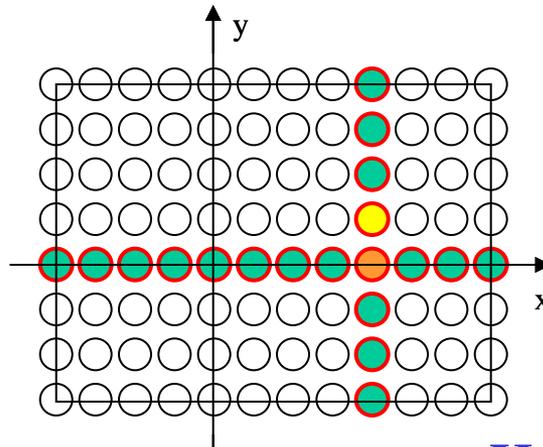
Appariement : stratégies d'optimisation

Le principe des stratégies d'optimisation est de parcourir « intelligemment » l'espace de recherche K pour minimiser la complexité de calcul de l'optimum d'appariement :



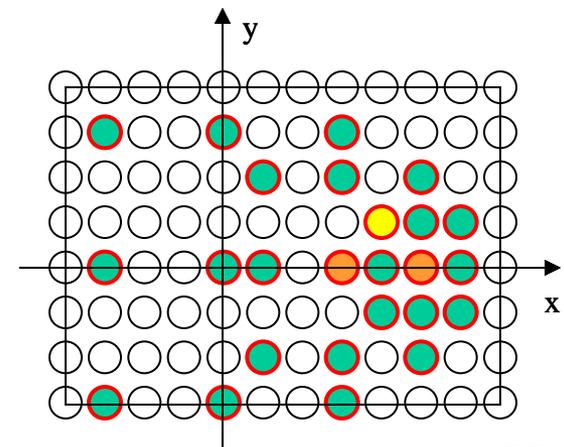
EXHAUSTIF

K



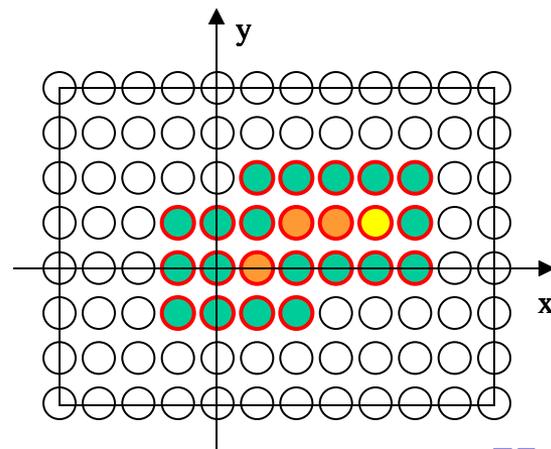
SEPARÉ

K



HIERARCHIQUE

K



DESCENTE DE GRADIENT

K

Pour chaque point de K , on ne calcule la somme des différences que tant qu'on demeure inférieur à un seuil T . Le meilleur déplacement correspond à celui pour lequel le plus de points ont été examinés.

SEQUENTIEL

- Ces différentes techniques ne sont pas exclusives, et peuvent être combinées.

- En l'absence d'hypothèse sur la fonction d'appariement (séparabilité, convexité...) dans le domaine K , seul le parcours exhaustif garantit le calcul d'un optimum *global* sur K .

Exemple : normes mpeg de codage video



Flot optique sous-résolu

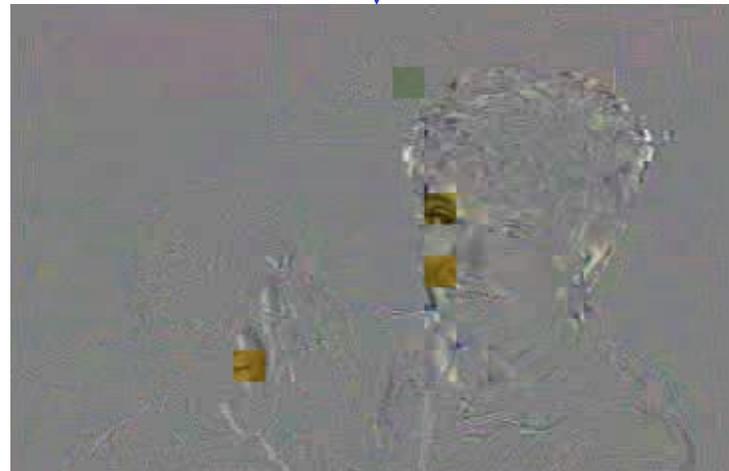


Trame prédite

- Découpage en blocs 16×16 ou 32×32 .
- Hypothèse de même déplacement pour tous les pixels du blocs.
- Codage : déplacements + erreurs comprimées.

↳ Compression statique *JPEG* :

- Transformée en cosinus discrète
- Quantification
- Codage entropique



Erreurs de prédiction

Référence :

VcDemo
TU Delft

Techniques différentielles (1)

LUCAS & KANADE 1981

Le principe de la méthode de Lucas et Kanade est de calculer le minimum de la fonction d'appariement quadratique (SSD), en supposant que le déplacement recherché est petit, comme le point où les dérivées de la fonction d'appariement s'annulent, par rapport à δ_x et à δ_y .

$$\begin{aligned}(v_x^t, v_y^t) &= \arg \min_{(\delta x, \delta y) \in K} A(\delta x, \delta y) \\ &= \arg \min_{(\delta x, \delta y) \in K} \sum_{(x, y) \in B} (I(x, y, t) - I(x + \delta x, y + \delta y, t + 1))^2\end{aligned}$$

Sous l'hypothèse que I est régulière et que le déplacement (δ_x, δ_y) est petit, on écrit le développement de Taylor à l'ordre 1 de I :

$$I(x + \delta x, y + \delta y, t + 1) \approx I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t}$$

La fonction d'appariement A devient :

$$A(\delta x, \delta y) \approx \sum_{(x, y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right)^2$$

Techniques différentielles (1)

LUCAS & KANADE 1981

$$A(\delta x, \delta y) \approx \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right)^2$$

Equations d'Euler-Lagrange de minimisation de A : Annulation des dérivées premières par rapport à δx et à δy :

$$2 \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right) \frac{\partial I}{\partial x} = 0$$

$$2 \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right) \frac{\partial I}{\partial y} = 0$$

Ce qui revient à la résolution du système linéaire :

$$\mathbf{H} \cdot \mathbf{v} = \mathbf{b}$$

$$\text{avec } \mathbf{H} = \begin{pmatrix} \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \right)^2 & \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial y} \right)^2 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} - \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial t} \\ - \sum_{(x,y) \in B} \frac{\partial I}{\partial y} \frac{\partial I}{\partial t} \end{pmatrix}$$

et $\mathbf{v} = (v_x, v_y)$ le déplacement recherché.

La résolution du système (S) est finalement réalisée par une méthode itérative, type Gauss-Seidel :

$$\text{Initialisation : } v_x^{(0)} = 0; v_y^{(0)} = 0$$

$$\text{Pour } k > 0 : \begin{cases} v_x^{(k)} = \frac{1}{\mathbf{H}_1^1} (\mathbf{b} - \mathbf{H}_1^2 \cdot v_y^{(k-1)}) \\ v_y^{(k)} = \frac{1}{\mathbf{H}_2^2} (\mathbf{b} - \mathbf{H}_2^1 \cdot v_x^{(k)}) \end{cases}$$

Techniques différentielles (1)

LUCAS & KANADE 1981

L'existence et la stabilité d'une solution au système (S) dépend de la matrice \mathbf{H} :

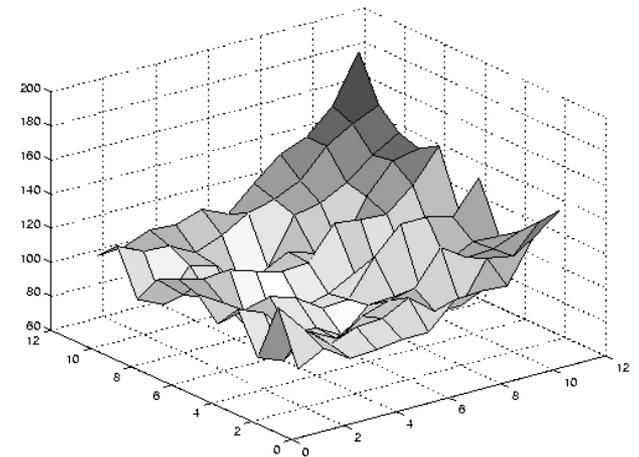
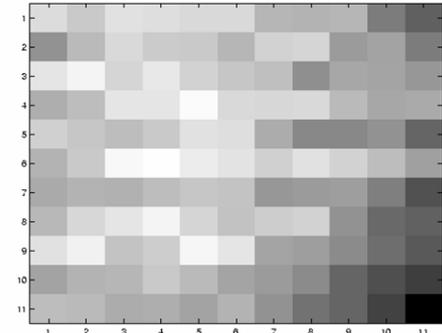
$$\mathbf{H} = \begin{pmatrix} \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x}\right)^2 & \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial y}\right)^2 \end{pmatrix}$$

NB : On retrouve la matrice de structure utilisée pour le calcul des points de Harris...

La matrice \mathbf{H} doit être de rang 2 « au sens fort », c'est-à-dire posséder 2 valeurs propres (λ_1, λ_2) *grandes*. Les auteurs proposent d'utiliser $\lambda = \min(\lambda_1, \lambda_2)$ comme indicateur de confiance du déplacement calculé.

Cette propriété correspond à une interprétation algébrique du *problème de l'ouverture*...

Rang 0 : zone homogène

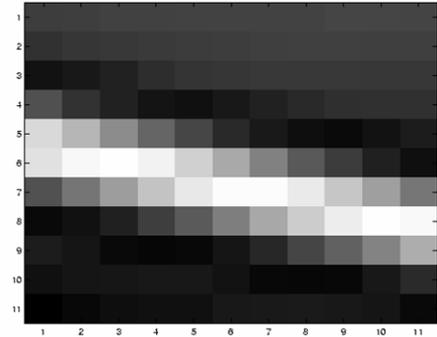


Zone faiblement texturée :
2 valeurs propres faibles

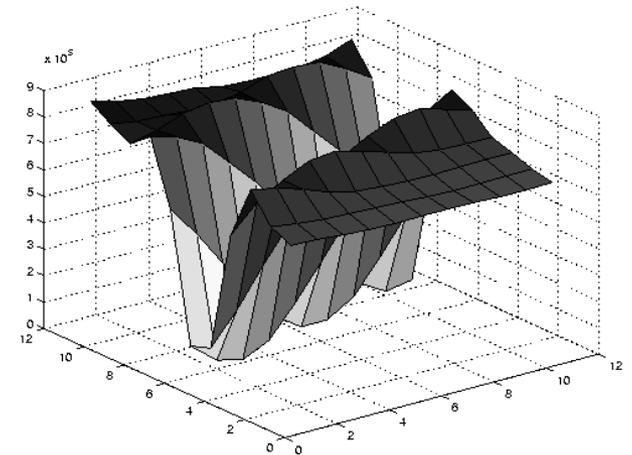
Sources : Steven Seitz UW-Seattle

Fonction d'appariement A

Rang 1 : contour simple



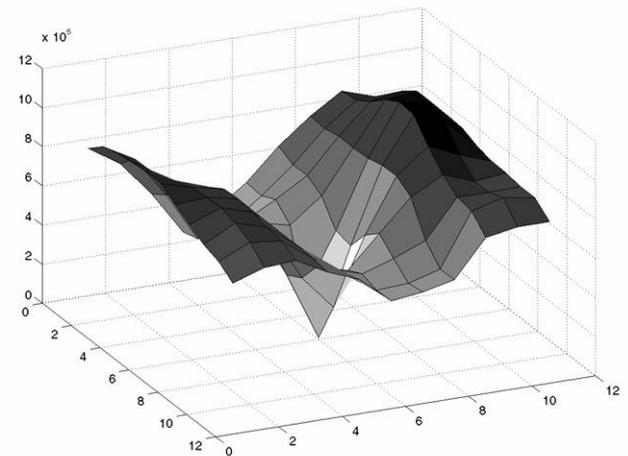
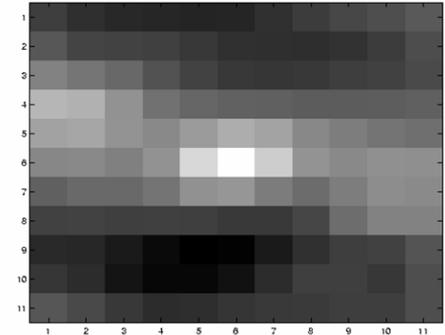
Contour rectiligne : 1 seule
grande valeur propre



Sources : Steven Seitz UW-Seattle

Fonction d'appariement A

Rang 2 : point anguleux



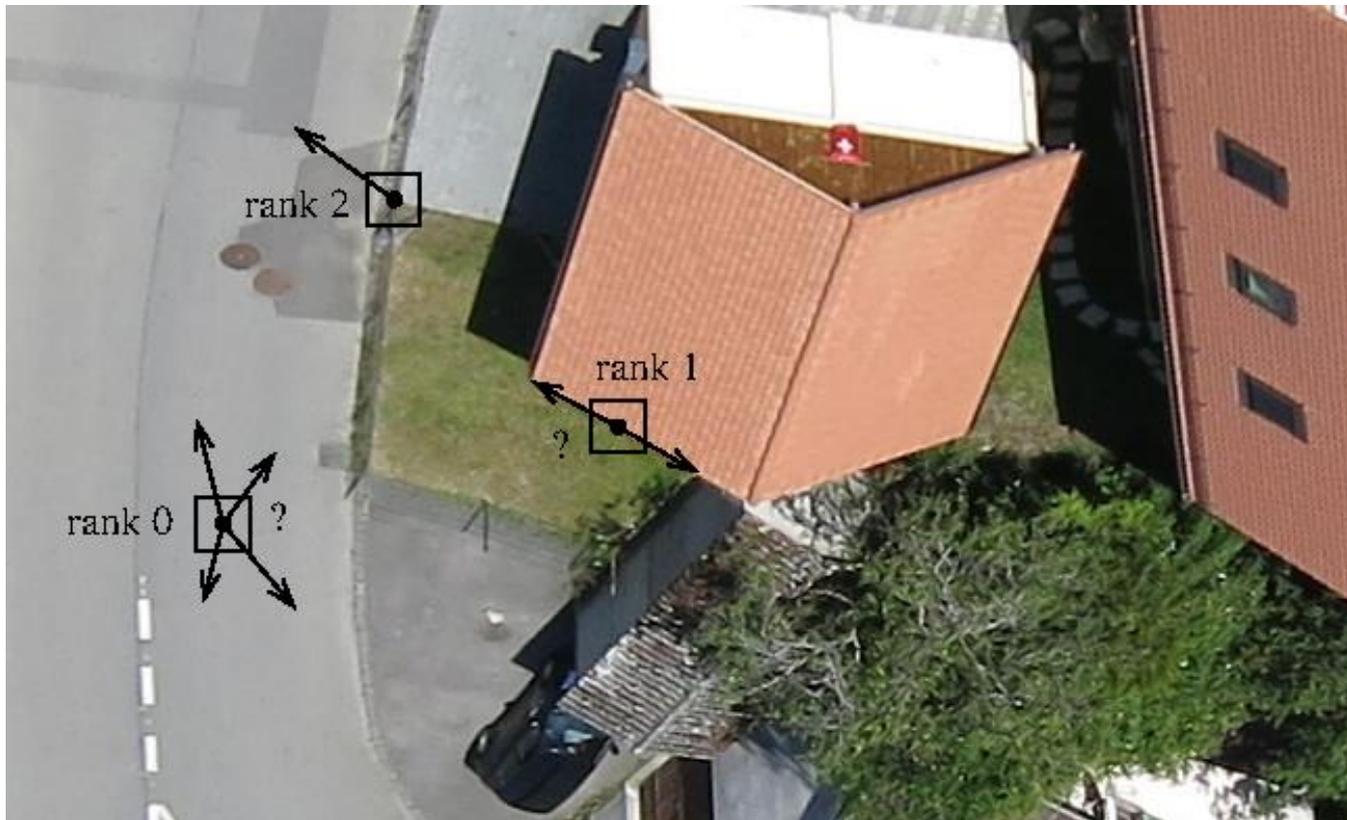
Point anguleux : 2 valeurs
propres importantes

Sources : Steven Seitz UW-Seattle

Fonction d'appariement A

Flot Optique L&K épars

En pratique donc, la méthode de Lucas & Kanade ne fournit des estimations fiables que lorsque la matrice de structure \mathbf{H} est de rang 2, c'est-à-dire pour les points coins !



Flot Optique L&K épars

- Le calcul étant indépendant pour chaque pixel, la méthode LK peut être massivement parallélisée.
- Mais en raison de la limitation liée à la matrice de structure, elle ne fournit qu'un flot épars, pas un champ dense.



Flot optique LK pyramidal (OpenCV)

Techniques différentielles (2)

HORN & SCHUNCK 1981

Exploitation directe de la contrainte (1) :

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

Formule de Taylor

$$= I(x, y, t) + \frac{\partial I}{\partial x} \cdot \delta x + \frac{\partial I}{\partial y} \cdot \delta y + \frac{\partial I}{\partial t} \cdot \delta t + \dots$$

Termes d'ordres supérieures négligés

SOIT :

$$\nabla I \cdot \mathbf{v} + \frac{\partial I}{\partial t} = 0$$

avec

$$\left\{ \begin{array}{l} \nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \quad \text{gradient spatial} \\ \mathbf{v} = (v_x^t, v_y^t) \quad \text{inconnues} \\ \frac{\partial I}{\partial t} \quad \text{gradient temporel} \end{array} \right.$$

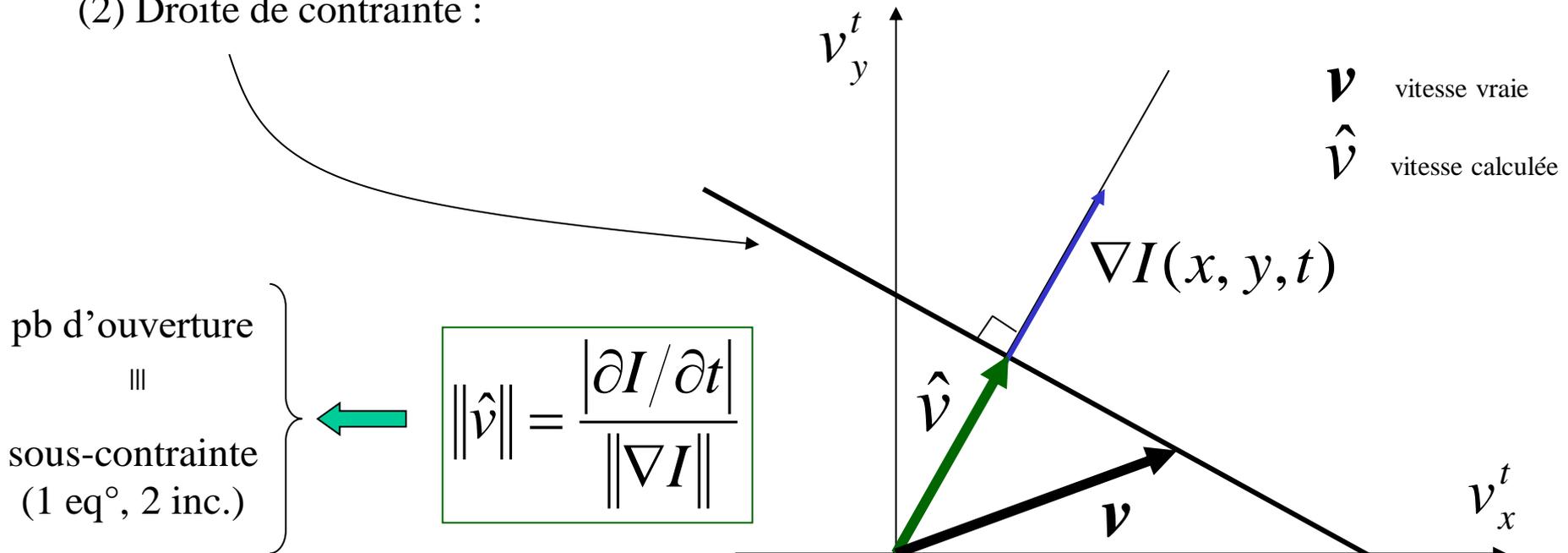
Equation de contrainte du mouvement apparent (ECMA)

ou : équation du flot optique

Interprétation de l'ECMA

(1) En faisant l'hypothèse d'une certaine *régularité* du champ et de *petits déplacements*, les *changements temporels* dans l'image sont *équivalents* (au premier ordre) au produit scalaire des *changements spatiaux* et de la *vitesse apparente*.

(2) Droite de contrainte :



Résolution de l'ECMA

HORN & SCHUNCK 1981

Résolution de l'ECMA par ajout d'une contrainte de régularité.

⇔ Régularisation du pb mal posé par hypothèse de champ lisse de déplacement.

→ Minimisation d'une fonction de coût :

$$C_{(x,y)}^t(v_x^t, v_y^t) = \underbrace{\left(\nabla I \cdot v + \frac{\partial I}{\partial t} \right)^2}_{\text{ECMA}} + \lambda \underbrace{\left[\left(\frac{\partial v_x^t}{\partial x} \right)^2 + \left(\frac{\partial v_x^t}{\partial y} \right)^2 + \left(\frac{\partial v_y^t}{\partial x} \right)^2 + \left(\frac{\partial v_y^t}{\partial y} \right)^2 \right]}_{\text{REGULARISATION}}$$

facteur de
pondération

Résolution de l'ECMA

HORN & SCHUNCK 1981

1. Minimisation d'une fonctionnelle quadratique :

$$C_{(x,y)}(u, v) = (I_x u + I_y v + I_t)^2 + \lambda(u_x^2 + u_y^2 + v_x^2 + v_y^2)$$

$$\text{avec : } \begin{cases} (u, v) \text{ composantes du champ (inconnues) à calculer} \\ \{I_x, I_y, I_t\} \text{ dérivées partielles de l'image} \\ \{u_x, u_y, v_x, v_y\} \text{ dérivées partielles des composantes du champ} \end{cases}$$

—————→ Annulation des dérivées premières $\partial/\partial u(\dots) = 0$; $\partial/\partial v(\dots) = 0$

Equations d'Euler-Lagrange de minimisation de C :

$$\begin{cases} 2(I_x u + I_y v + I_t)I_x + 2\lambda \left[\left(\frac{\partial}{\partial u} \frac{\partial u}{\partial x} \right) \frac{\partial u}{\partial x} + \left(\frac{\partial}{\partial u} \frac{\partial u}{\partial y} \right) \frac{\partial u}{\partial y} \right] = 0 \\ 2(I_x u + I_y v + I_t)I_y + 2\lambda \left[\left(\frac{\partial}{\partial v} \frac{\partial v}{\partial x} \right) \frac{\partial v}{\partial x} + \left(\frac{\partial}{\partial v} \frac{\partial v}{\partial y} \right) \frac{\partial v}{\partial y} \right] = 0 \end{cases}$$

$$\text{soit : } \begin{cases} (I_x u + I_y v + I_t)I_x + \lambda \Delta u = 0 \\ (I_x u + I_y v + I_t)I_y + \lambda \Delta v = 0 \end{cases} \quad \text{avec : } \begin{cases} \Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \\ \Delta v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \end{cases} \quad \text{laplaciens de } u \text{ et } v$$

Résolution de l'ECMA

2. Approximation du Laplacien : $\nabla^2 f = f - \tilde{f}$ \tilde{f} = moyenne de f dans un certain voisinage

soit :
$$\begin{cases} (v - \tilde{v})(I_x^2 + I_y^2 + \lambda) = -I_y(I_x \tilde{u} + I_y \tilde{v} + I_t) \\ (u - \tilde{u})(I_x^2 + I_y^2 + \lambda) = -I_x(I_x \tilde{u} + I_y \tilde{v} + I_t) \end{cases}$$

3. Schéma itératif de résolution :

Méthode de Gauss-Seidel :
$$\begin{cases} u^k = \tilde{u}^{k-1} - I_x(I_x \tilde{u}^{k-1} + I_y \tilde{v}^{k-1} + I_t) \cdot (I_x^2 + I_y^2 + \lambda)^{-1} \\ v^k = \tilde{v}^{k-1} - I_y(I_x \tilde{u}^{k-1} + I_y \tilde{v}^{k-1} + I_t) \cdot (I_x^2 + I_y^2 + \lambda)^{-1} \end{cases}$$

Soit, en reprenant les notations originales :

ALGORITHME DE HORN & SCHUNCK

avec :

$$\begin{cases} N = \frac{\partial I}{\partial x} \tilde{v}_x^{k-1} + \frac{\partial I}{\partial y} \tilde{v}_y^{k-1} + \frac{\partial I}{\partial t} \\ D = \lambda + \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 \end{cases}$$

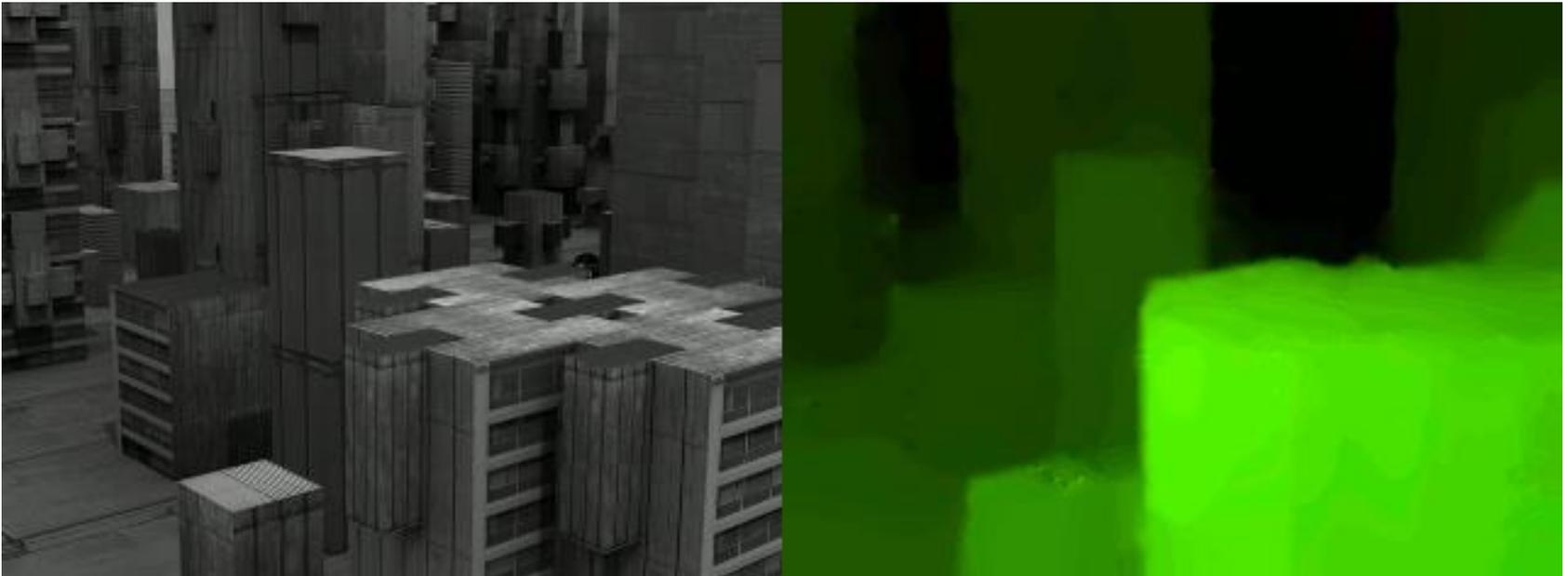
• Initialisation : $v_x^0 = 0$ $v_y^0 = 0$

• Répéter jusqu'à convergence :

$$\begin{cases} v_x^k = \tilde{v}_x^{k-1} - \frac{\partial I}{\partial x} \frac{N}{D} \\ v_y^k = \tilde{v}_y^{k-1} - \frac{\partial I}{\partial y} \frac{N}{D} \end{cases}$$

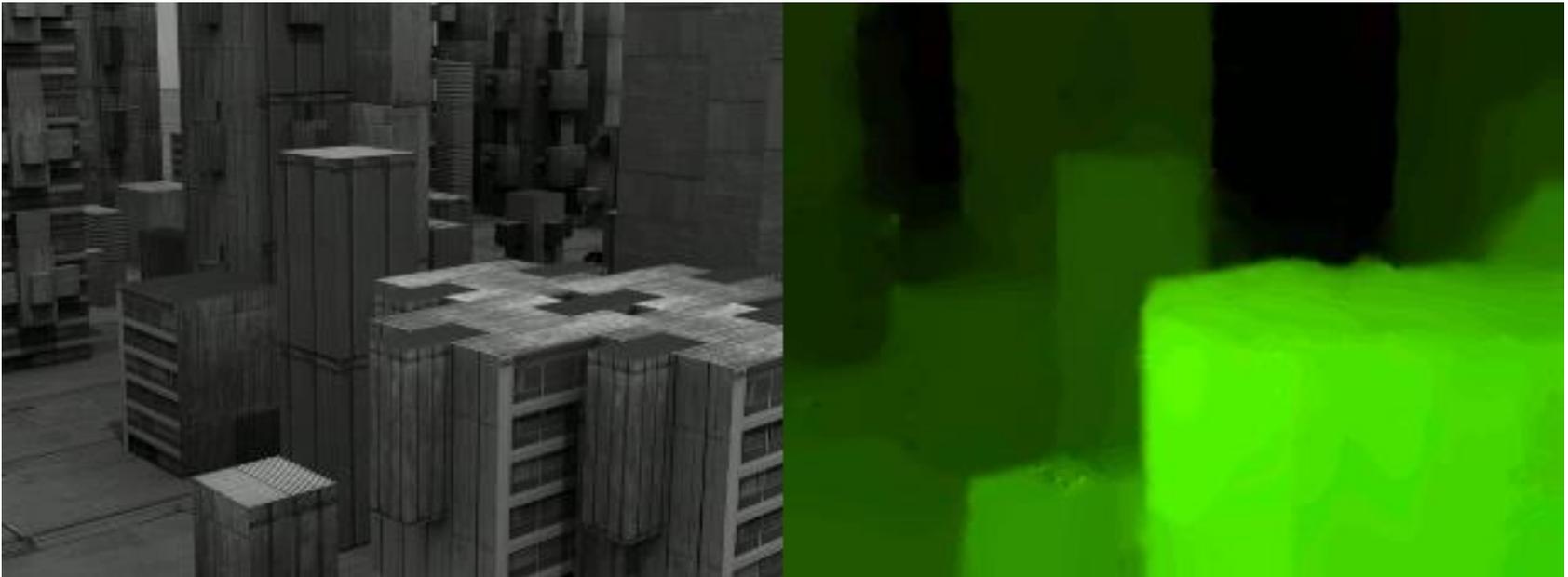
Flot Optique dense H&S

- A cause de l'estimation itérative des moyennes locales des composantes du flot optique, la méthode de Horn & Schunck est une méthode globale, qui ne peut pas être parallélisée aussi massivement que la méthode de Lucas & Kanade.
- Mais grâce à la régularisation spatiale explicite, elle fournit un flot optique dense.



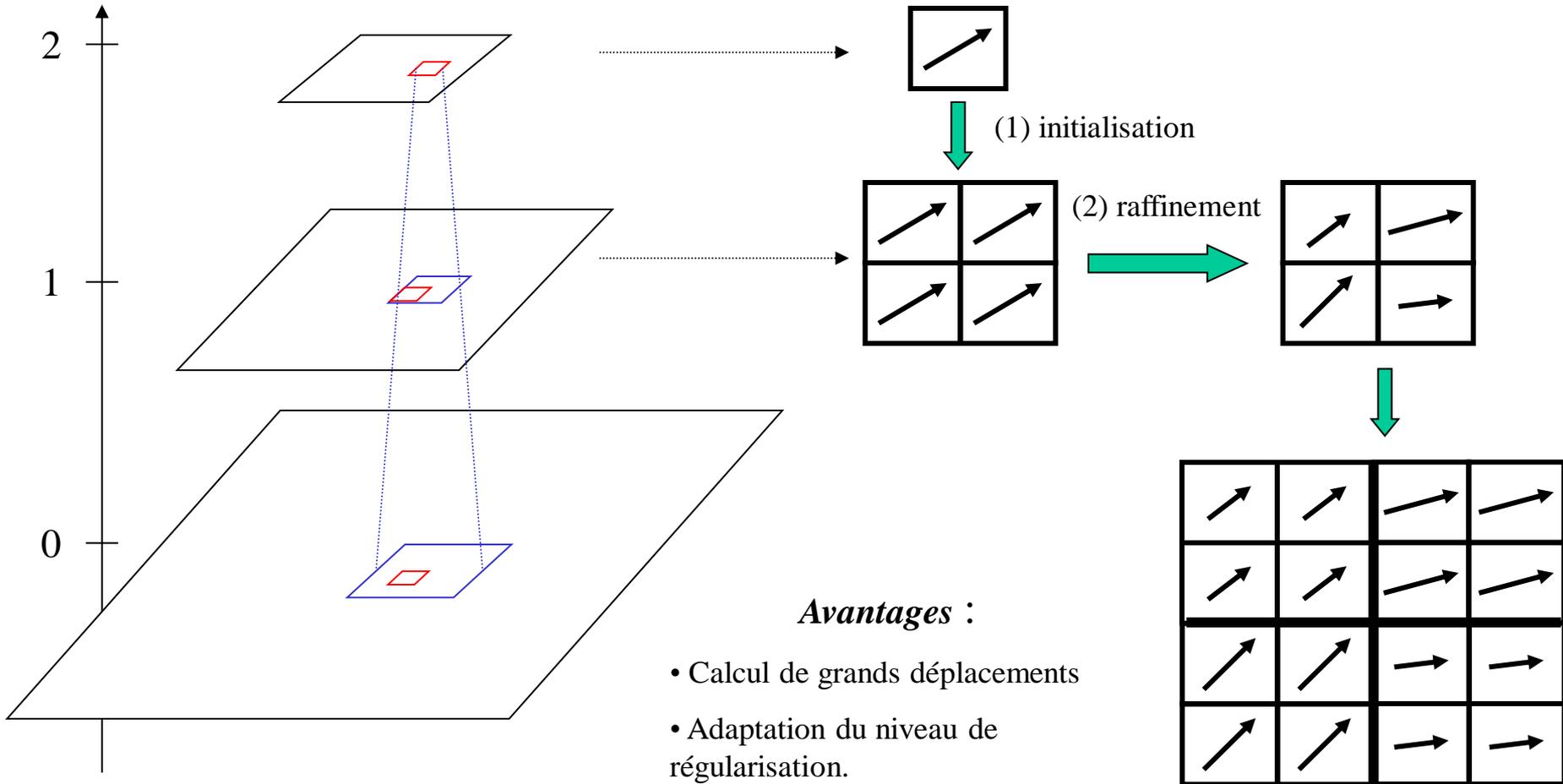
Flot Optique dense H&S

- A cause de l'estimation itérative des moyennes locales des composantes du flot optique, la méthode de Horn & Schunck est une méthode globale, qui ne peut pas être parallélisée aussi massivement que la méthode de Lucas & Kanade.
- Mais grâce à la régularisation spatiale explicite, elle fournit un flot optique dense.



Estimation multi-échelles

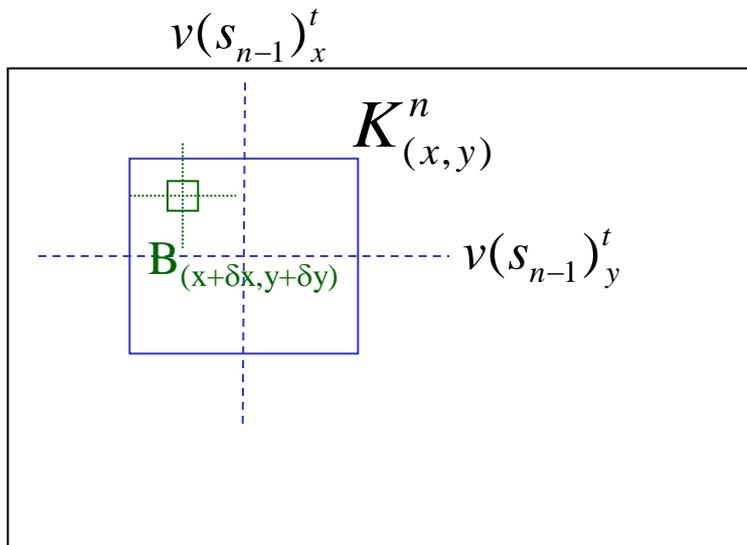
Niveau



Principe du calcul multi-échelles

ALGORITHME PAR CORRELATION

$$(v(s_n)_x^t, v(s_n)_y^t) = \arg \max_{(\delta x, \delta y) \in K^n} C_{(x,y)}^t(\delta x, \delta y)$$



avec : $K_{(x,y)}^n = [v(s_{n-1})_x^t \pm \Delta_1] \times [v(s_{n-1})_y^t \pm \Delta_2]$

ALGORITHME DIFFERENTIEL

- Initialisation : $v(s_0)_x^0 = 0$ $v(s_0)_y^0 = 0$
- Répéter jusqu'à convergence :

$$\left| \begin{array}{l} v(s_0)_x^k = \tilde{v}(s_0)_x^{k-1} - \frac{\partial I}{\partial x} \frac{N}{D} \\ v(s_0)_y^k = \tilde{v}(s_0)_y^{k-1} - \frac{\partial I}{\partial y} \frac{N}{D} \end{array} \right.$$

- Init. : $v(s_1)_x^0 = v(s_0)_x^\infty$ $v(s_1)_y^0 = v(s_0)_y^\infty$
- Répéter jusqu'à convergence :

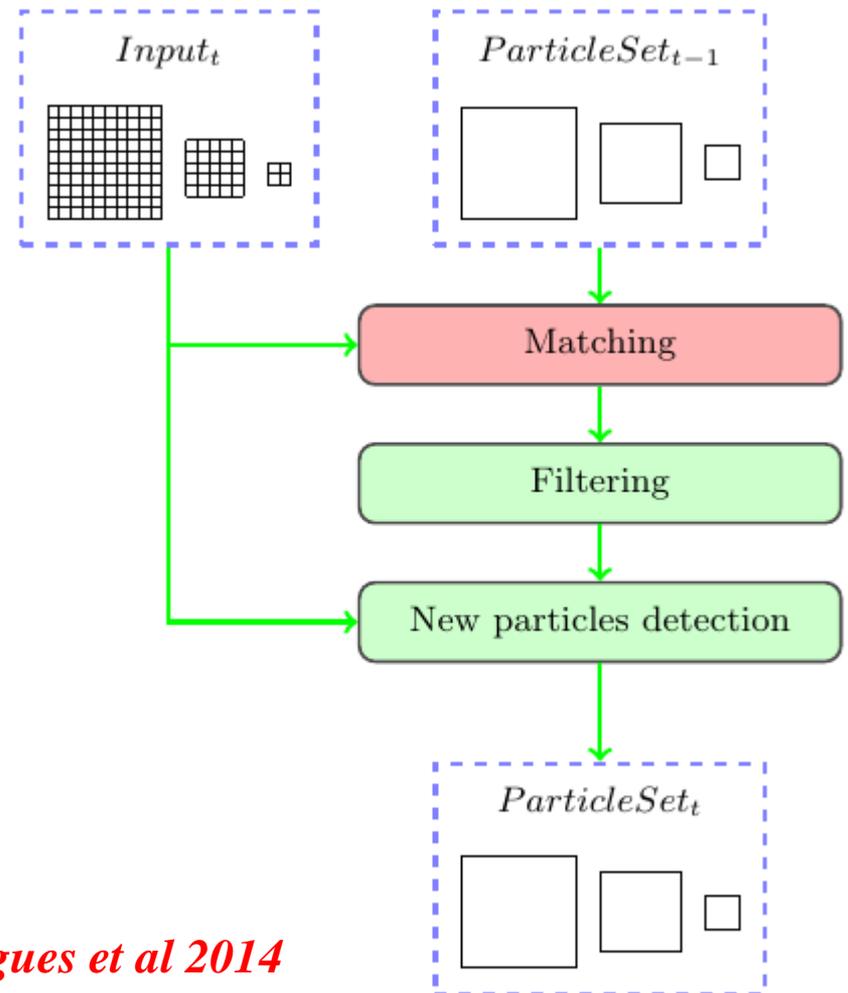
$$\left| \begin{array}{l} v(s_1)_x^k = \tilde{v}(s_1)_x^{k-1} - \frac{\partial I}{\partial x} \frac{N}{D} \\ v(s_1)_y^k = \tilde{v}(s_1)_y^{k-1} - \frac{\partial I}{\partial y} \frac{N}{D} \end{array} \right.$$

.../...

Suivi semi-dense long-terme

Video Extruder [Garrigues et al 2014] :

- Gestion massivement parallèle d'un champ de particules et leurs trajectoires.
- Prédiction hybride temporelle (trajectoires) et multi-échelles
- Produit un champ semi-dense de trajectoires.



Garrigues et al 2014

Suivi semi-dense long-terme



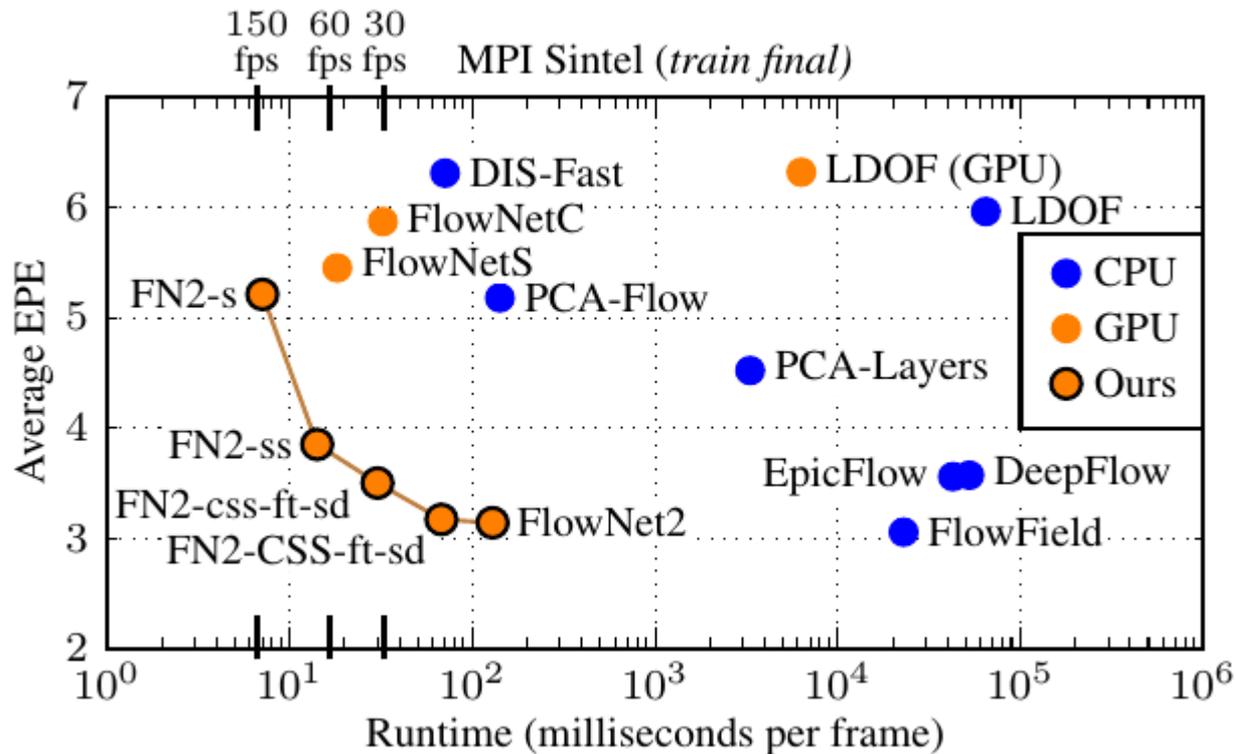
Garrigues et al 2014

Conclusion sur les méthodes analytiques

- Depuis les années 80 jusqu'à aujourd'hui, l'estimation du flot optique est resté un domaine de recherche très actif.
- Il a donné lieu à des milliers de publications, qui reposent sur un tout petit nombre de principes de base, mais énormément d'astuces et de recettes différentes...
- Quelques bases de données annotées existent pour l'évaluation :
 - <https://vision.middlebury.edu/flow/> (Intérieur, dense)
 - http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=flow (Extérieur, épars)

Méthodes par apprentissage

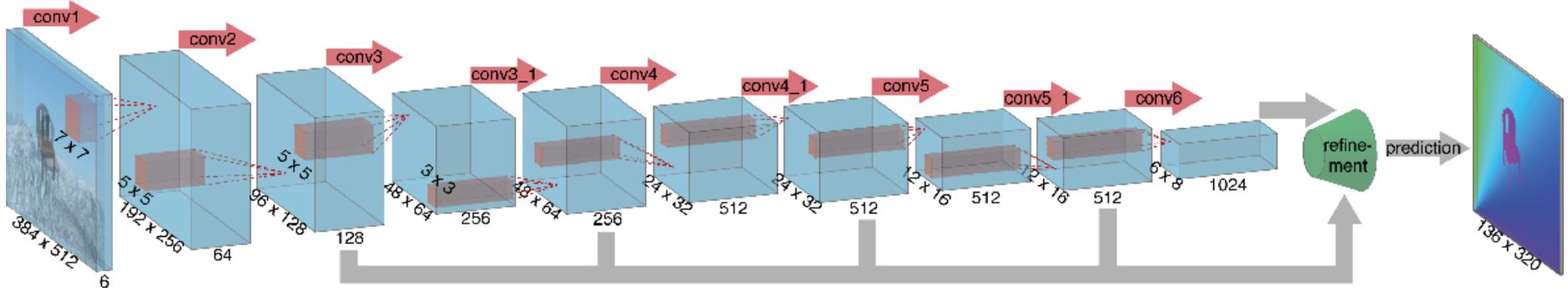
Les méthodes end-to-end fondées sur le Deep Learning émergent massivement depuis 2015, avec des résultats en progression foudroyante, que ce soit en termes de précision ou d'efficacité computationnelle :



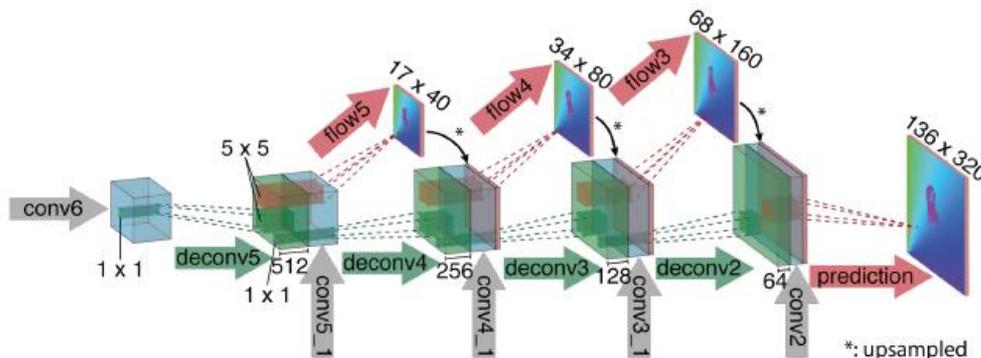
[Ilg et al 2017]

Méthodes par apprentissage

FlowNetSimple



Réseau CNN FlowNetSimple pour l'estimation du flot optique [Fischer 2015]

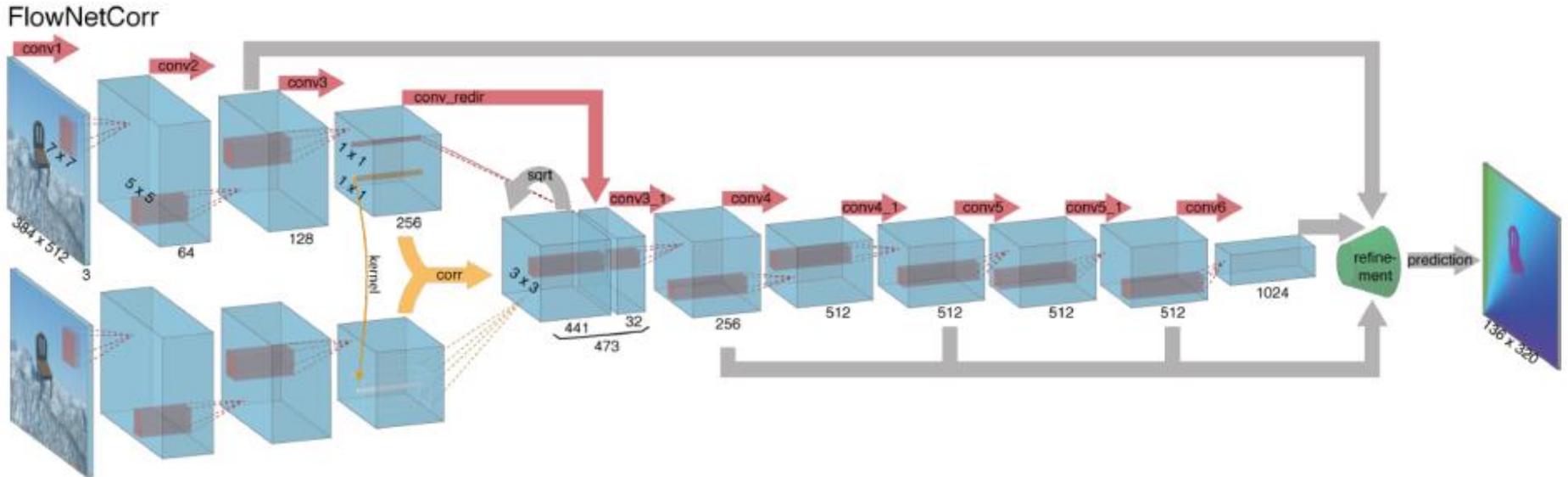


Sous-réseau décodeur (raffinement) [Fischer 2015]

- Le réseau profond réalise une prédiction dense en exploitant l'ensemble des indices visuels.
- Avec des annotations denses, la fonction de loss peut être tout simplement la distance L_2 :

$$\|v_{GT} - v_E\|_2$$

Méthodes par apprentissage



Réseau CNN FlowNetCorr pour l'estimation du flot optique **[Fischer 2015]**

- Contrairement à FlowNetSimple, FlowNetCorr apprend les caractéristiques spatiales des images fixes, qui sont corrélées explicitement dans le bloc 4d « corr » (non appris) qui est fourni en entrée d'un sous-réseau encodeur.
- La distinction FlowNetSimple / FlowNetCorr renvoie au pré-traitement ou à la sélection spatiale des méthodes analytiques.

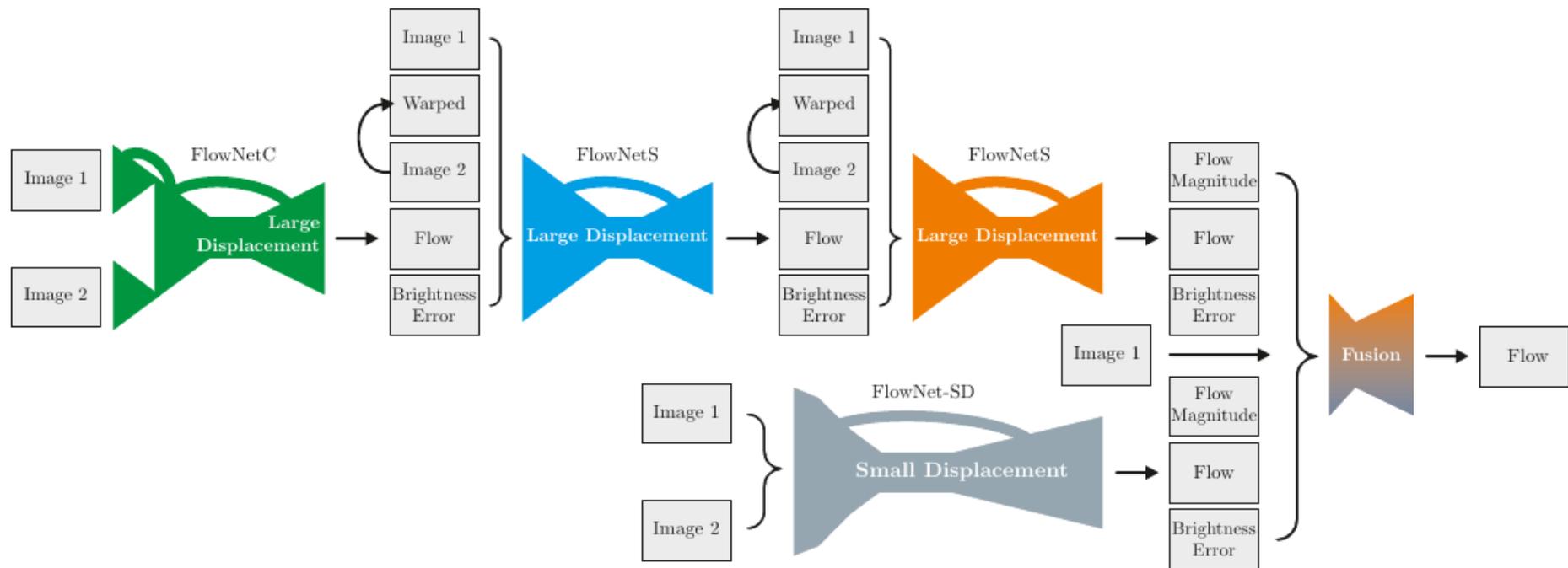
Quelles bases d'apprentissage ?



La base de données semi-synthétique *Flying Chairs* [**Fischer 2015**] : annotations denses faciles à augmenter (à gauche).

Méthodes par apprentissage – FlowNet 2.0

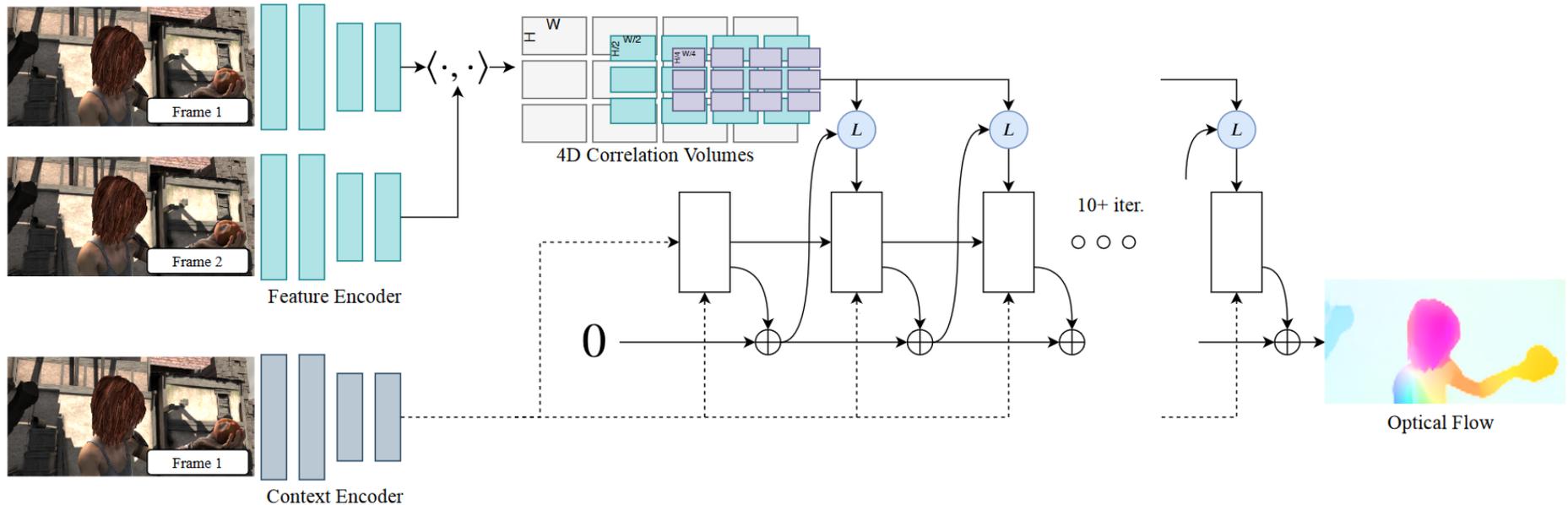
La complémentarité des performances entre FlowNetSimple et FlowNetCorr, ainsi que certaines limitations, comme la possibilité de gérer des petits *et* des grands déplacements, ont favorisé une tendance modulaire dans un certain nombre d'architectures :



FlowNet 2.0 [Ilg et al 2017]

Méthodes par apprentissage – RAFT

Le réseau RAFT (Recurrent All-Pairs Field Transforms) imite de façon plus directe les méthodes analytiques, non seulement en reprenant le principe de corrélation intégrale des features via le volume 4d de corrélation, mais aussi en mimant le principe itératif des méthodes d'optimisation, par l'utilisation de blocs récurrents (GRUs) :



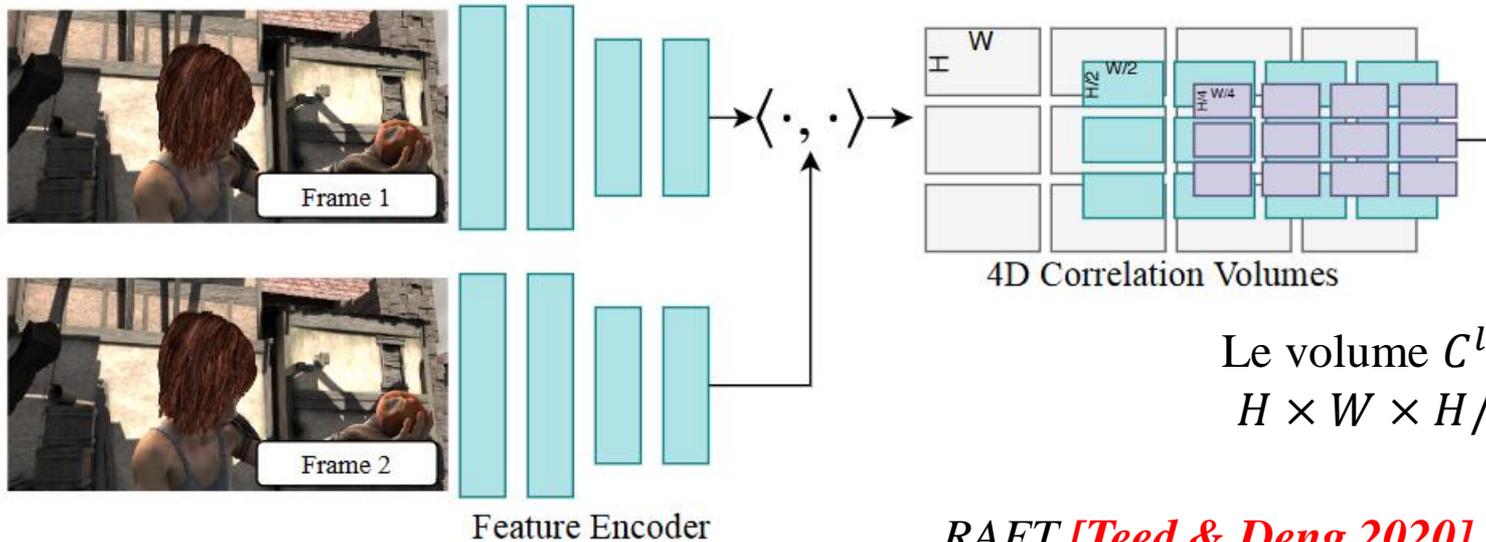
RAFT [Teed & Deng 2020]

Méthodes par apprentissage – RAFT

La première phase de RAFT consiste à calculer – à l’instar de FlowNetCorr – une carte de caractéristiques (feature maps) pour chaque image de la paire, puis à corrélérer les 2 cartes obtenues dans un volume 4d :

$$C(x, y, x', y') = F_1(x, y) \cdot F_2(x', y') = \sum_i F_1^i(x, y) F_2^i(x', y')$$

RAFT forme ensuite une pyramide de volumes 4d de corrélation $\{C^0, C^1, C^2, C^3\}$ où seules les 2 dernières dimensions sont quantifiées.



Le volume C^l est de taille $H \times W \times H/2^l \times W/2^l$

RAFT [Teed & Deng 2020]

Méthodes par apprentissage – RAFT

La seconde phase de RAFT consiste en l'application récursive d'une cellule GRU (Gated Recurrent Unit) convolutionnelle, qui estime itérativement (indice d'itération : k) le flot résiduel Δf et l'ajoute à l'estimé courant : $f_k = f_{k-1} + \Delta f$, en fonction de l'état « mémoire » m_k , de l'état « entrée » i_k , et de l'état « caché » h_k , à partir duquel on calcule Δf . W_m , W_i et W_h forment les poids appris des convolutions :

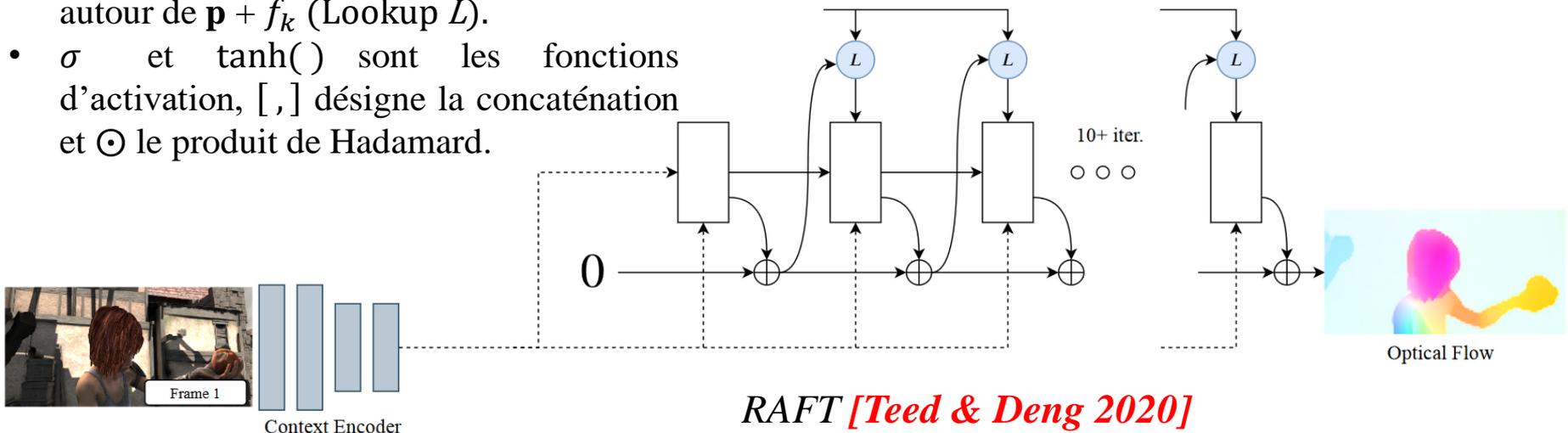
$$m_k = \sigma(W_m * [h_{k-1}, x_k])$$

$$i_k = \sigma(W_i * [h_{k-1}, x_k])$$

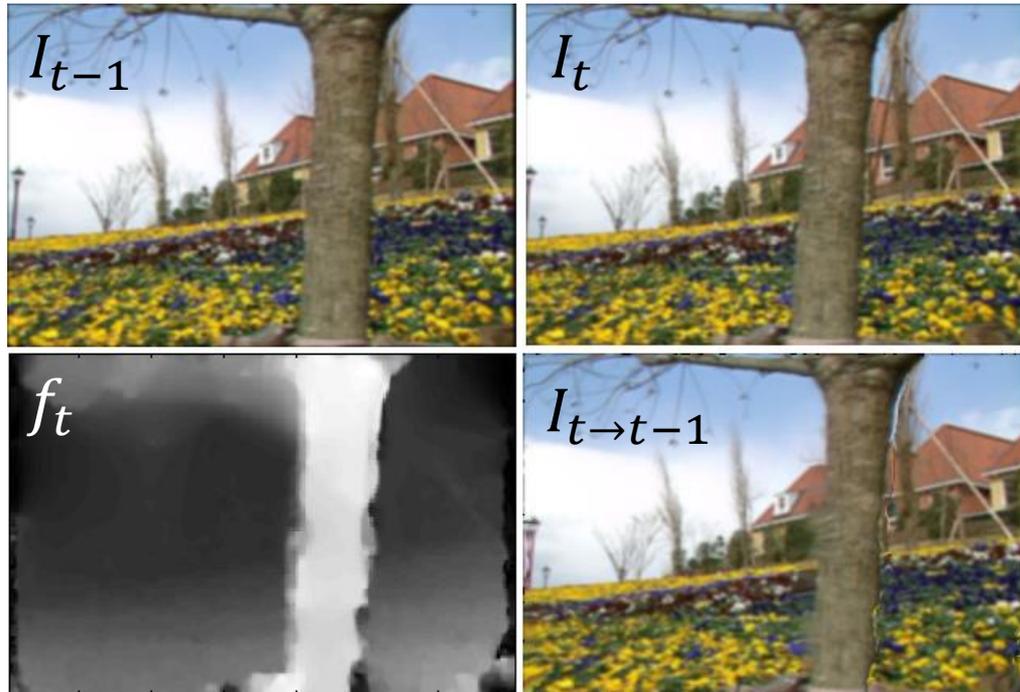
$$\hat{h}_k = \tanh(W_h * [i_k \odot h_{k-1}, x_k])$$

$$h_k = (1 - m_k) \odot h_{k-1} + m_k \odot \hat{h}_k$$

- Le vecteur d'entrée x_k est formé, pour chaque pixel \mathbf{p} , de la concaténation de son flot estimé f_k , et des features de corrélation et de contexte associées, i.e. autour de $\mathbf{p} + f_k$ (Lookup L).
- σ et $\tanh(\cdot)$ sont les fonctions d'activation, $[,]$ désigne la concaténation et \odot le produit de Hadamard.



Apprentissage auto-supervisé ?



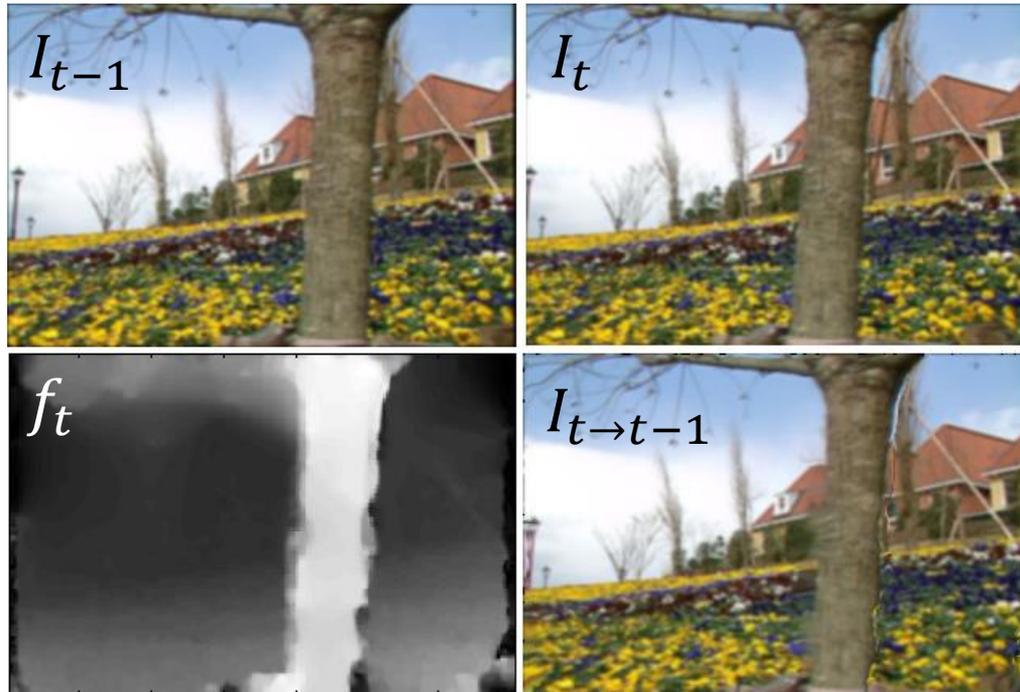
L'apprentissage (ou le fine tuning) auto-supervisé peut être fait sur des images naturelles en utilisant la fonction de *loss photométrique*, qui mesure la différence entre une image et sa prédiction fondée sur le flot optique.

$$\mathcal{L}_{ph} = \|I_{t-1} - I_{t \rightarrow t-1}\|$$

avec :

$$I_{t \rightarrow t-1}(\mathbf{p}) = I_t(\mathbf{p} + f_t(\mathbf{p}))$$

Apprentissage auto-supervisé ?



Des difficultés supplémentaires apparaissent néanmoins :

- Occultations
- Zones homogènes

Ces difficultés impliquent entre autres une modélisation plus fine de la fonction de coût (Loss), par ex. :

$$\mathcal{L}_{zh} = \|(I_{t-1} - I_{t \rightarrow t-1}) \cdot \|\nabla I_{t-1}\| \|\|$$

$$\mathcal{L}_{occ} = \min(\|I_{t-1} - I_{t \rightarrow t-1}\|, \|I_t - I_{t-1 \rightarrow t}\|)$$

Conclusion du chapitre

- ❑ EXPLOITATION DU FLOT OPTIQUE
 - ❑ Mouvement dominant, Profondeur, Temps avant collision...
- ❑ METHODES PAR APPARIEMENT
 - ❑ Compression Mpeg : OF sous résolu + Images résiduelles
- ❑ METHODES DIFFERENTIELLES
 - ❑ Lucas et Kanade, Horn et Schunck.
 - ❑ Pré-sélection spatiale vs Régularisation spatiale
- ❑ METHODES MULTI-ECHELLES
 - ❑ Continuités spatiale et temporelle par double prédiction
- ❑ METHODES PAR APPRENTISSAGE
 - ❑ Régularisation « naturelle », Exploite tous les indices visuels,...
 - ❑ Auto-supervision ?

Bibliographie (1)

- **B.D. Lucas & T. Kanade 1981** « *An iterative image registration technique with an application to stereo vision* » *International Journal of Computer Vision and Artificial Intelligence* 674-679
- **B.K.P Horn & B. Schunck 1981** « *Determining Optical Flow* » *Artificial Intelligence* 23 185-203
- **D.H. Ballard & C.M Brown 1982** « *Computer Vision* » Prentice Hall (Ch. 3, Ch. 7)
- **J.M. Jolion & A. Rosenfeld 1994** « *A pyramid framework for early vision* » Kluwer Academic Publishers Dordrecht, NL
- **R. Jain, R. Kasturi & B. Schunck 1995** « *Machine Vision* » McGraw-Hill Inc. (Ch.14)
- **J. Tautz, H.R. Heilmann & D.C Sandeman 2008** « **The Buzz about Bees: Biology of a Superorganism** ». Springer Berlin Heidelberg.
- **H.G.Krapp & M.Wicklein 2008** « **Central processing of visual information in insects** ». In: *The Senses: A Comprehensive Reference*, ed. by A.Basbaum, A.Kaneko, G.M.Shepherd, G.Westheimer, Academic Press, p.131–204.

Bibliographie (2)

- ***K.M. Steele 2014***, « **Psychology of Perception** », Lecture material, Appalachian State University, Fall 2014 <http://www1.appstate.edu/~kms/>
- ***M. Garrigues, A. Manzanera, & T.M. Bernard 2014*** ” **Video Extruder: a semi-dense point tracker for extracting beams of trajectories in real time**” *Journal of Real-Time Image Processing*. 2014. pp. 1-14.
- ***A. Dosovitskiy, P. Fischer et al 2015*** « **FlowNet: Learning Optical Flow with Convolutional Networks** » IEEE Int. Conf. on Computer Vision (ICCV), 2758-2766
- ***E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy & T. Brox 2017*** « **FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks** » IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- ***Z. Teed & J. Deng 2020*** « **RAFT: Recurrent All-Pairs Field Transforms for Optical Flow**» European Conf. on Computer Vision (ECCV'20), LNCS 12347, pp 402-419.