

MASTER
INFORMATIQUE



Nicolas Loménie

Vision par Ordinateur 3D en Mouvement

MASTER
INFORMATIQUE



Université
de Paris

Nicolas Loménie

Analyse de Séquences Vidéo

MOUVEMENT
MOUVEMENT

Objectifs :

- Présenter les concepts de base, l'importance et les problèmes liés à l'analyse du mouvement ;
- Comprendre les notions de champs de mouvement et de flot optique ainsi que leurs liens;
- Connaître l'équation d'invariance de la luminance;
- Savoir implémenter des méthodes de calcul de flot optique.
- Savoir implémenter des méthodes de tracking et apprendre la technique de filtrage de Kalman.

$I(x,y,t)$ →

$I(x,y,t+\Delta t)$ →

$I(x,y,t+2\Delta t)$ →



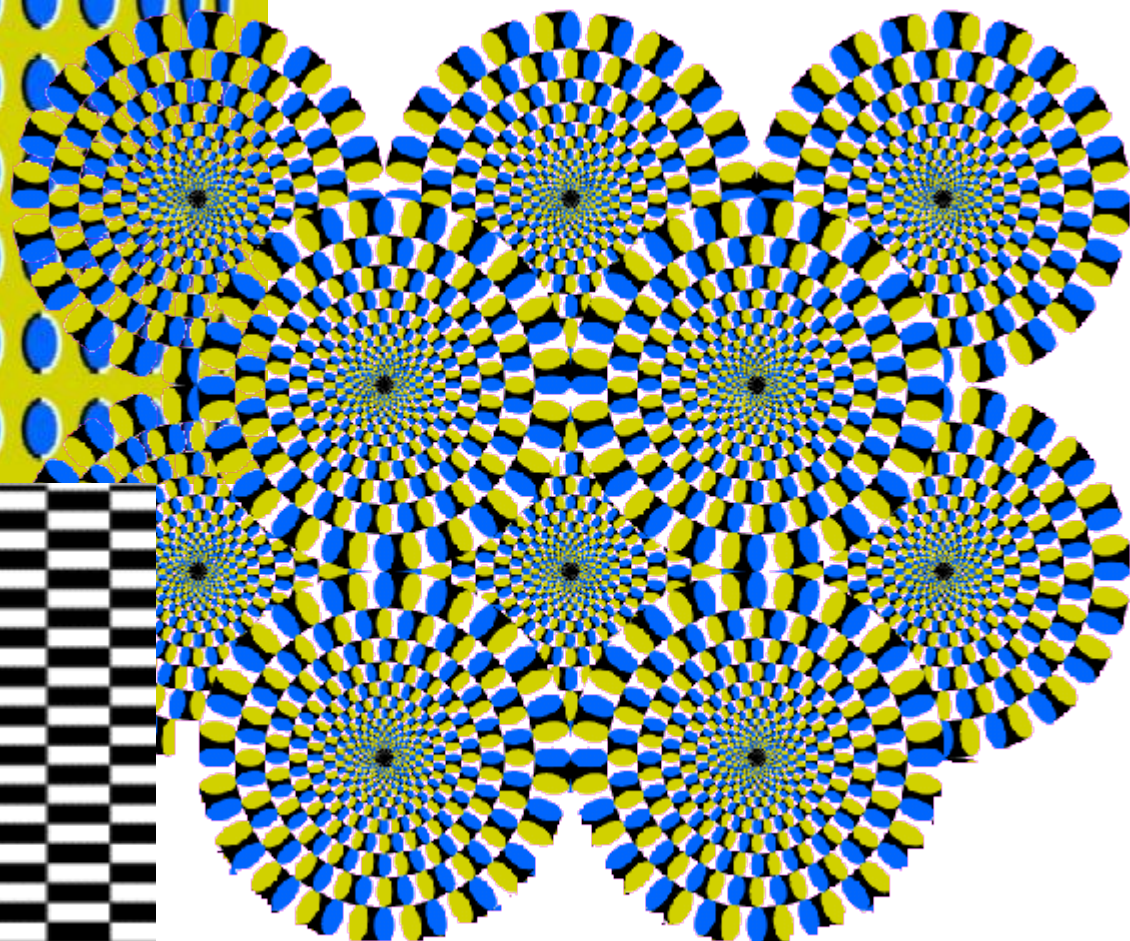
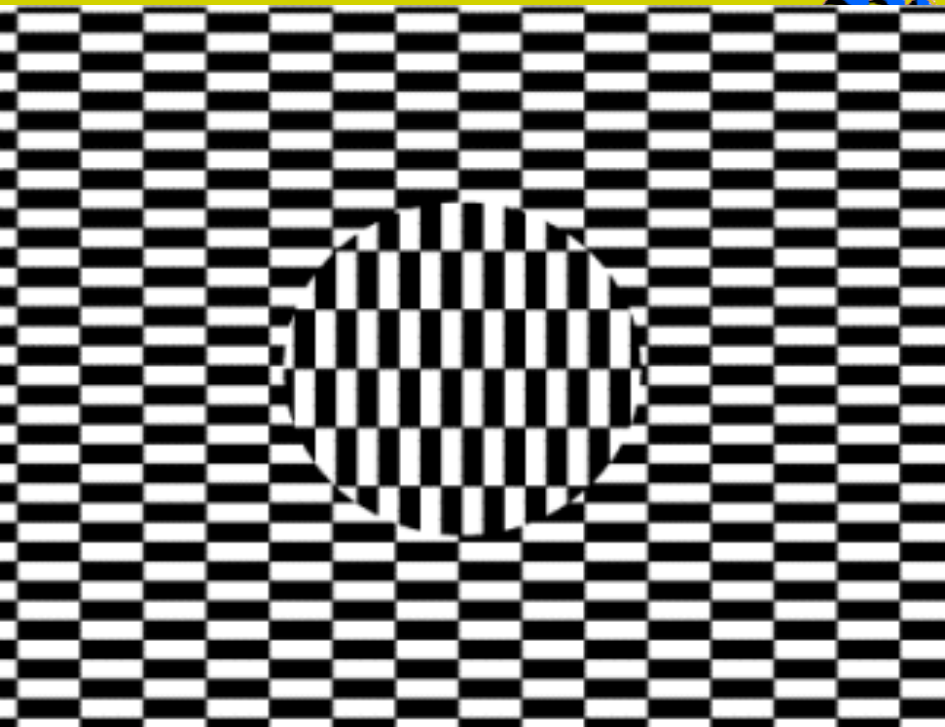
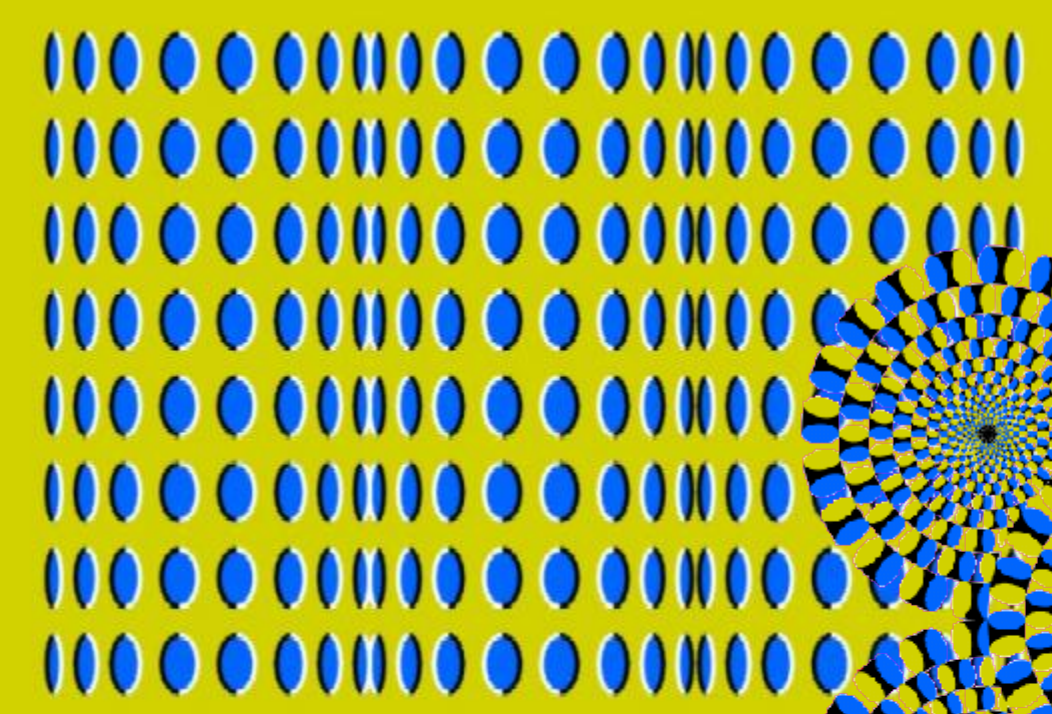
Introduction

Nous allons à présent nous occuper de l'information visuelle que l'on peut extraire à partir des changements spatio-temporels survenant dans une séquence d'images.

Séquence d'images

Une séquence d'images est une série de N images, ou frames, acquises à des instants du temps discret $t_k = t_0 + k\Delta t$, où Δt est un intervalle de temps fixe, et $k=0, 1, \dots, N-1$.

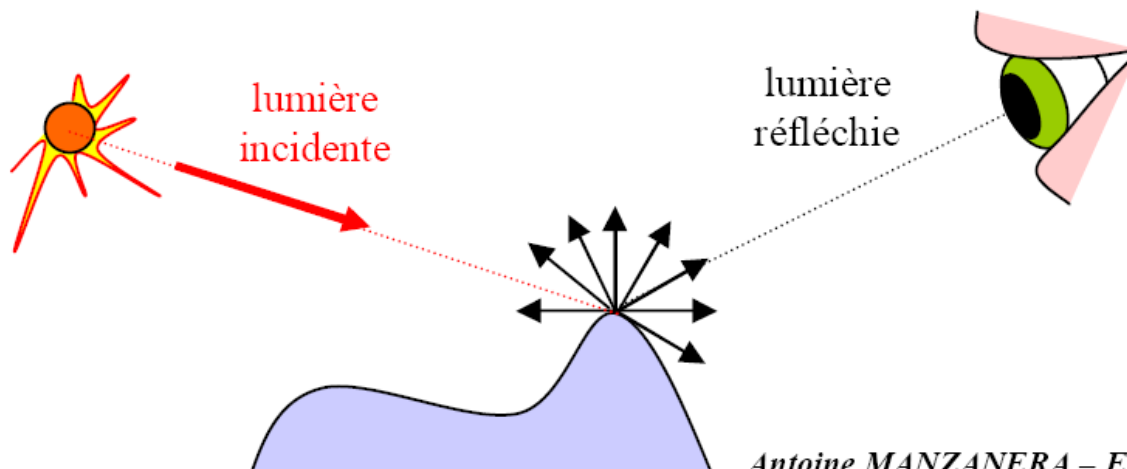
Remarque : nécessité d'un “frame grabber” capable de stocker des frames à des taux rapides. (“Frame rate” : $\Delta t = 1/24s$ et “field rate” : $\Delta t = 1/30s$). Il vous appartient de définir le Δt utile à votre problème et représentatif en terme d'échantillonnage temporel de la réalité des mouvements continus dans votre séquence. Persistence rétienne : $1/12s \rightarrow$ illusion vidéo ?



Hypothèse : MOUVEMENT \Leftrightarrow VARIATION d'INTENSITE

Si l'on suppose que les conditions d'illumination ne varient pas (\Leftarrow) et que les surfaces sont lambertiennes (\Rightarrow), les changements de luminance dans l'image sont dus à un mouvement relatif entre la caméra et la scène :

- Caméra mouvant devant une scène statique;
- Parties de la scène mouvante devant une caméra statique;
- Caméra et objets mouvant ensemble.



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

Importance du mouvement visuel

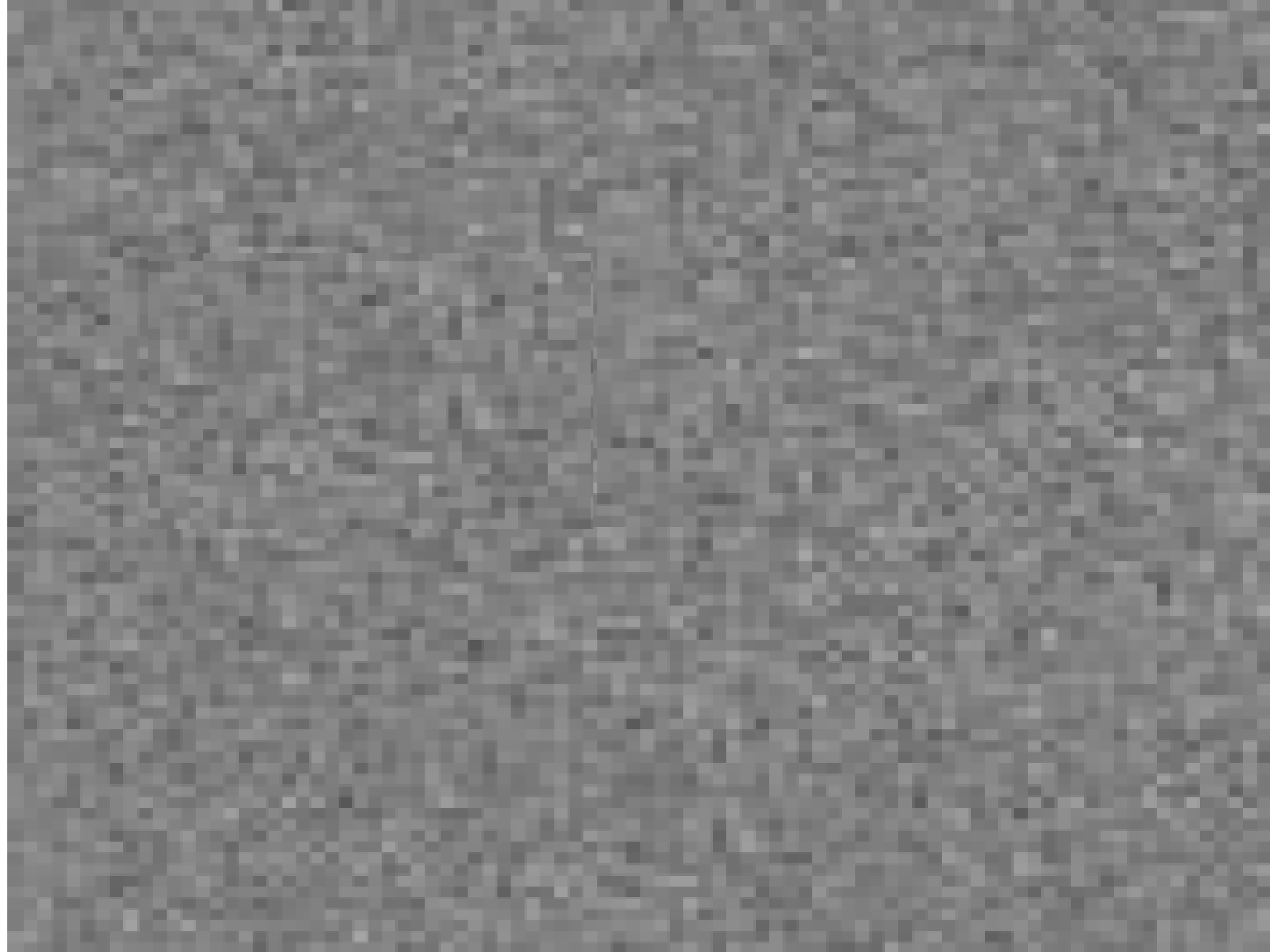
1. Le mouvement apparent d'objets sur le plan image est un indice visuel important pour comprendre la structure et le mouvement 3D dans une scène :

Structure 3D à partir du mouvement 2D ...

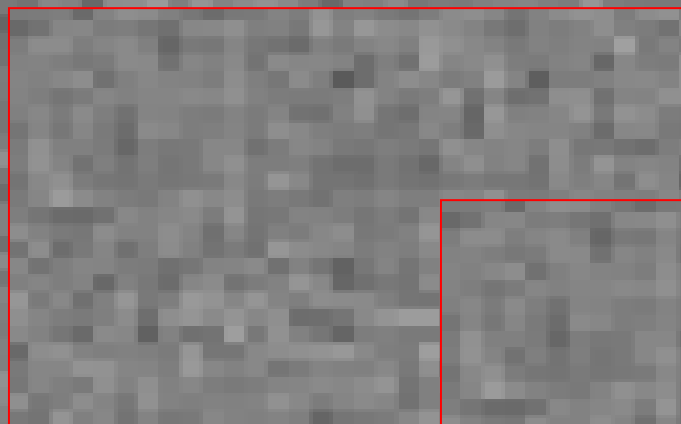
AnimatedModel

AnimatedPoint
Set

Ou forme





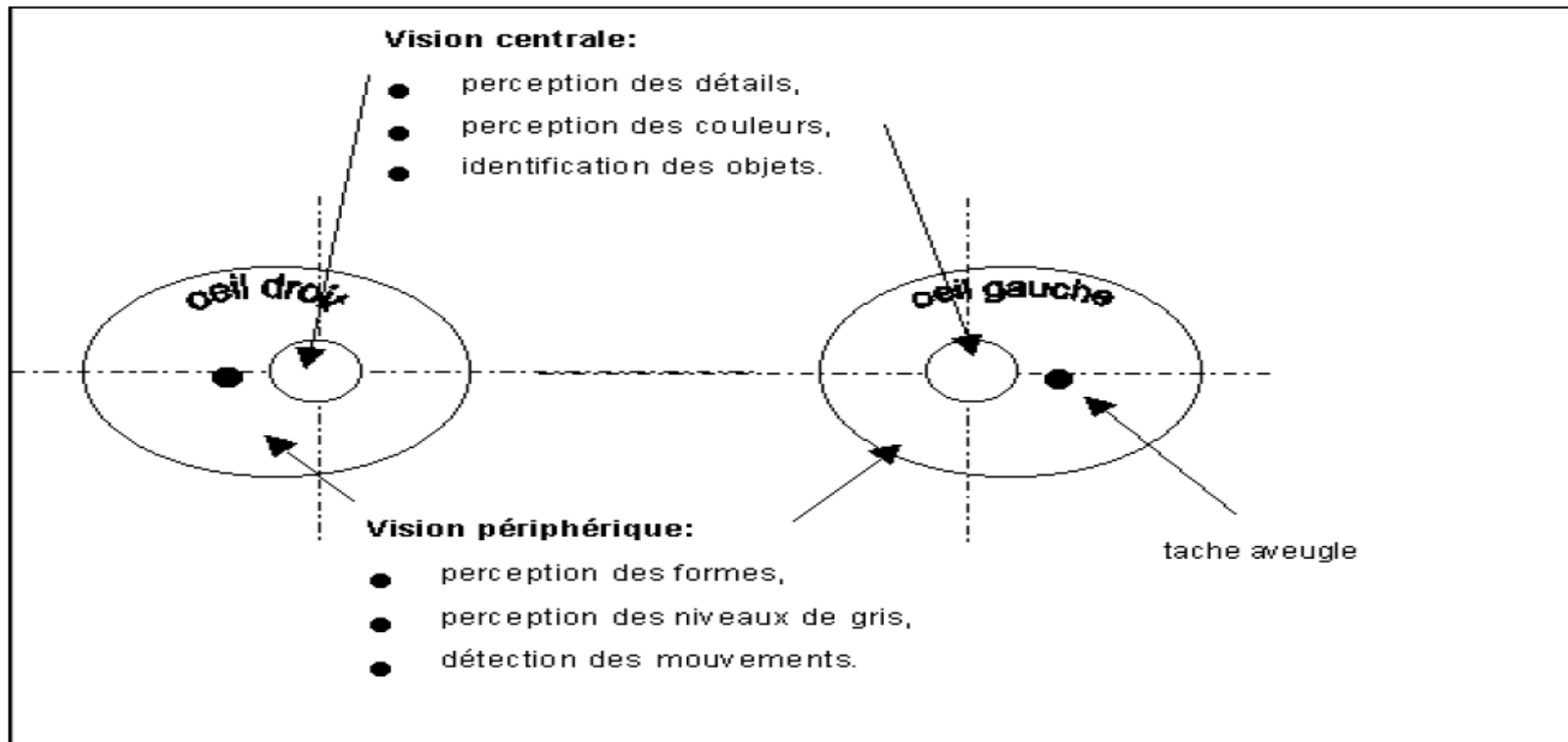


Importance du mouvement visuel

- Les systèmes biologiques utilisent le mouvement visuel pour inférer des informations sur le monde 3D avec peu de connaissance *a priori* sur celui-ci

<http://nivea.psycho.univ-paris5.fr/FeelingSupplements/BlindSpotFillingInExperiments.htm>

<https://www.verywellmind.com/why-do-you-have-a-blind-spot-2794880>



vision périphérique des vertébrés

Qu'est-ce qu'on peut inférer à partir du mouvement ?

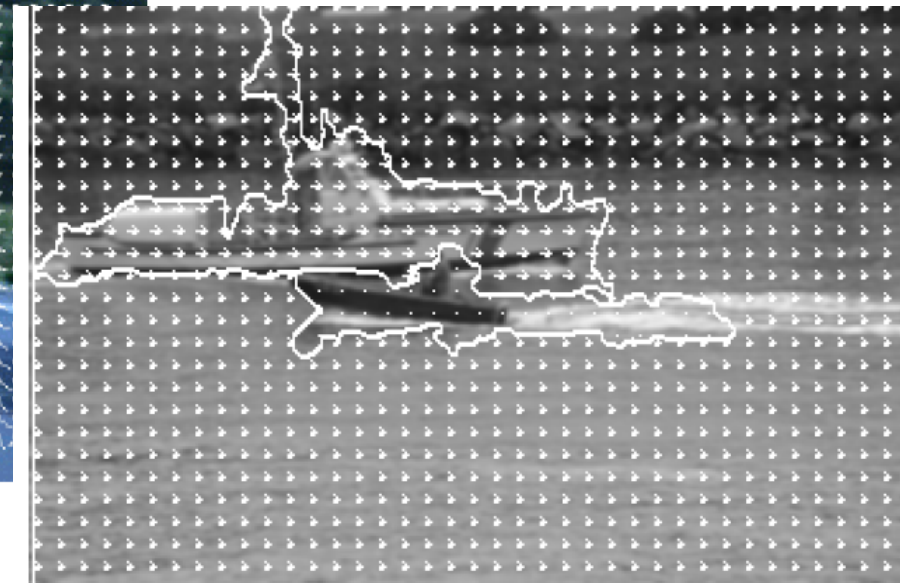
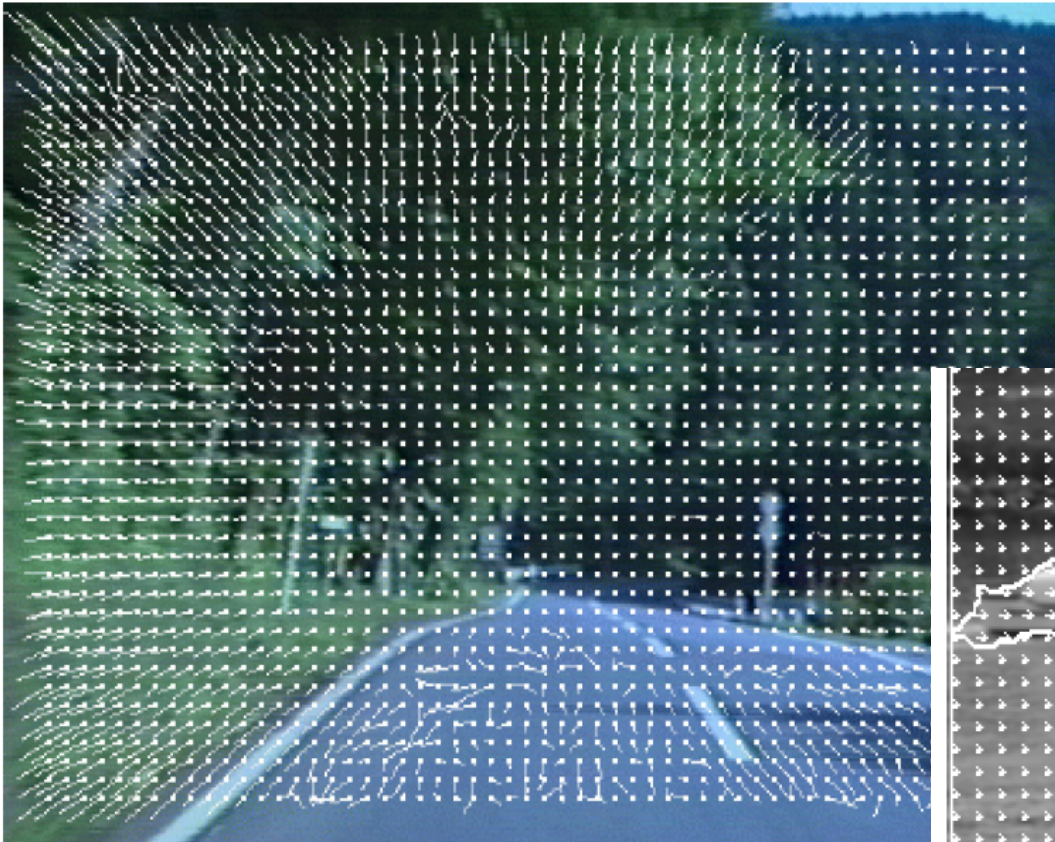
Détection de changement

Following Traffic



Analyse par Flot optique

Optical Flow due to camera motion



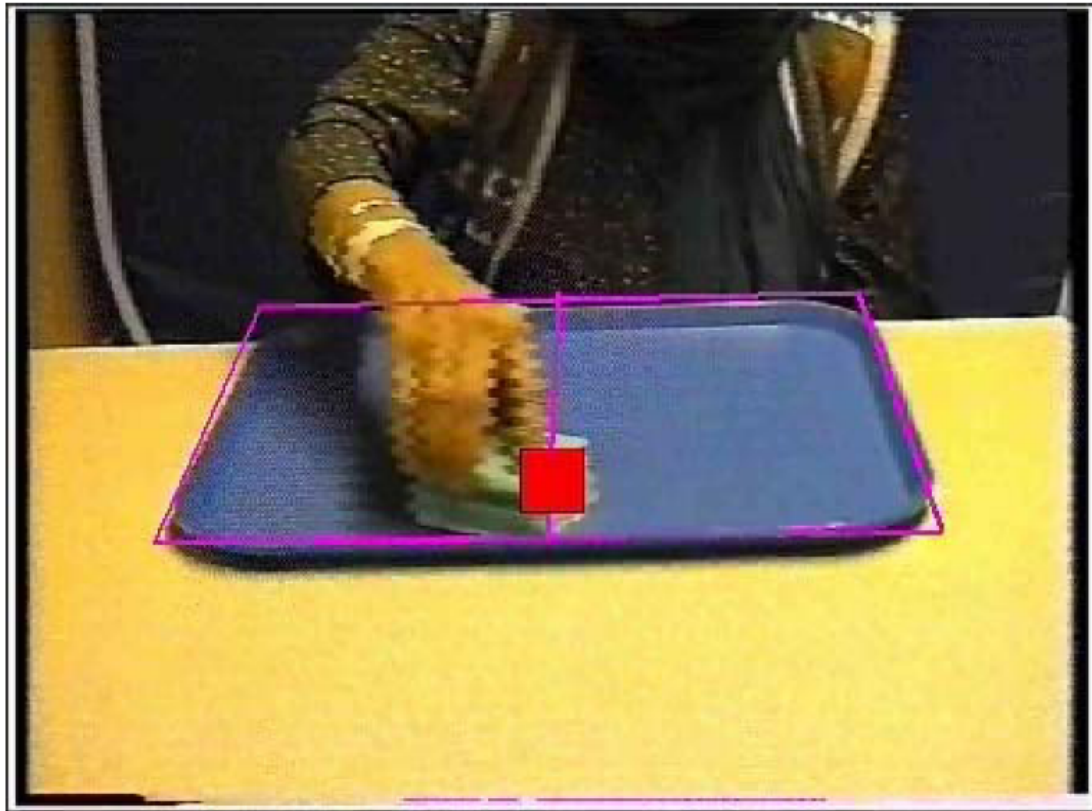
Motion-based segmentation.

Camera panning left (following smaller boat)

Larger boat moving right.

Tracking

Example: hand tracking



Measuring spatial neglect in stroke patients

Model-based tracking

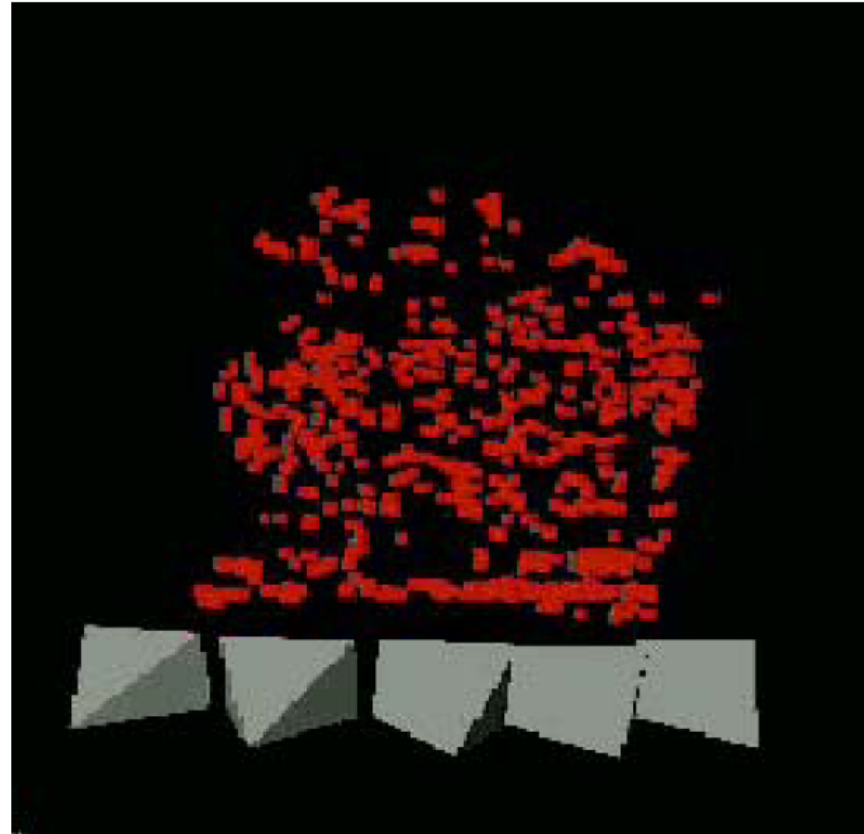
Feature-based Tracking

Chris Needham
and Roger Boyle,
University of
Leeds

Models based on colour
Wide variety of shapes
Multiple objects
Tracking through
occlusions

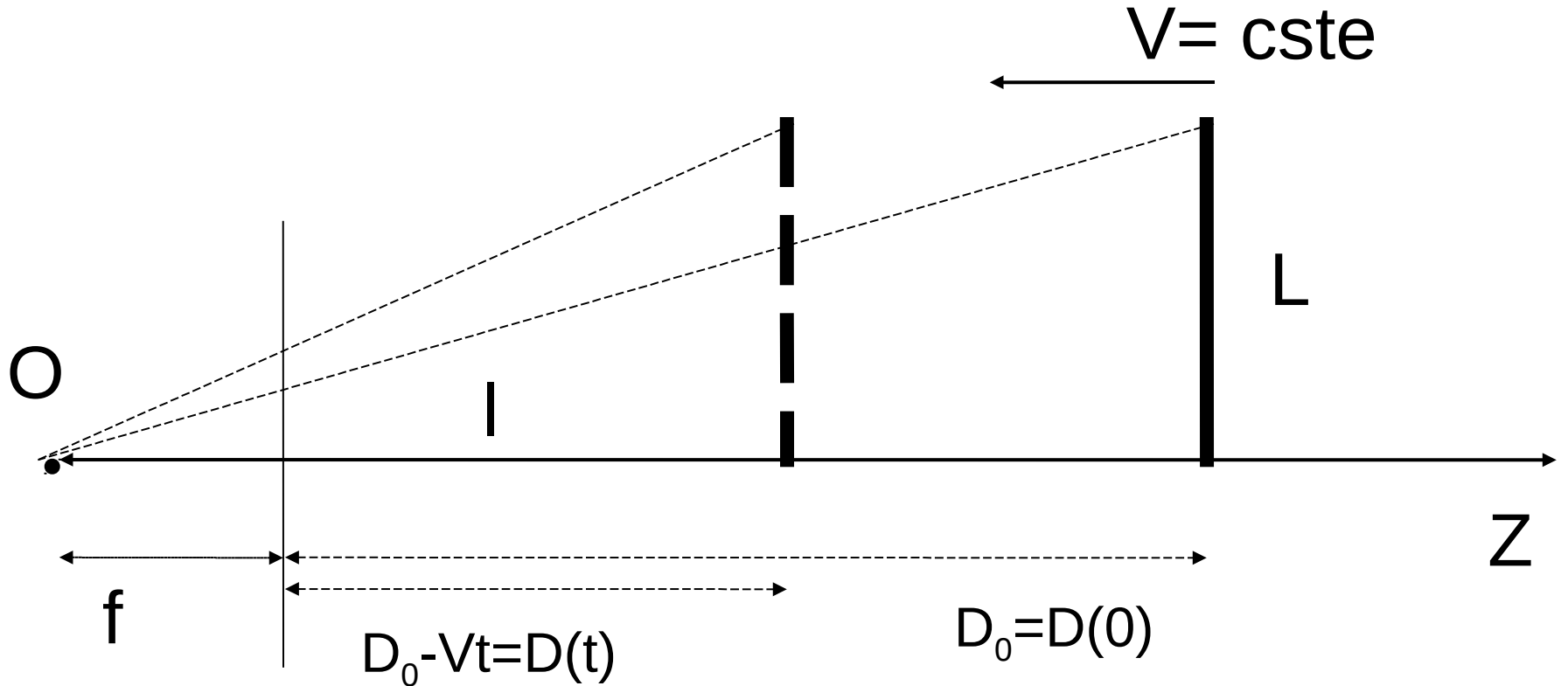


Structure From Motion



Time-to-Impact

Il est possible de calculer le temps, τ , nécessaire à la barre pour atteindre la caméra seulement à partir d'informations images. C'est-à-dire sans connaître la taille réelle de la barre ou sa vitesse dans l'espace 3D.



L, V, f et D_0 sont inconnus mais on peut écrire : $\tau = \frac{D}{V}$

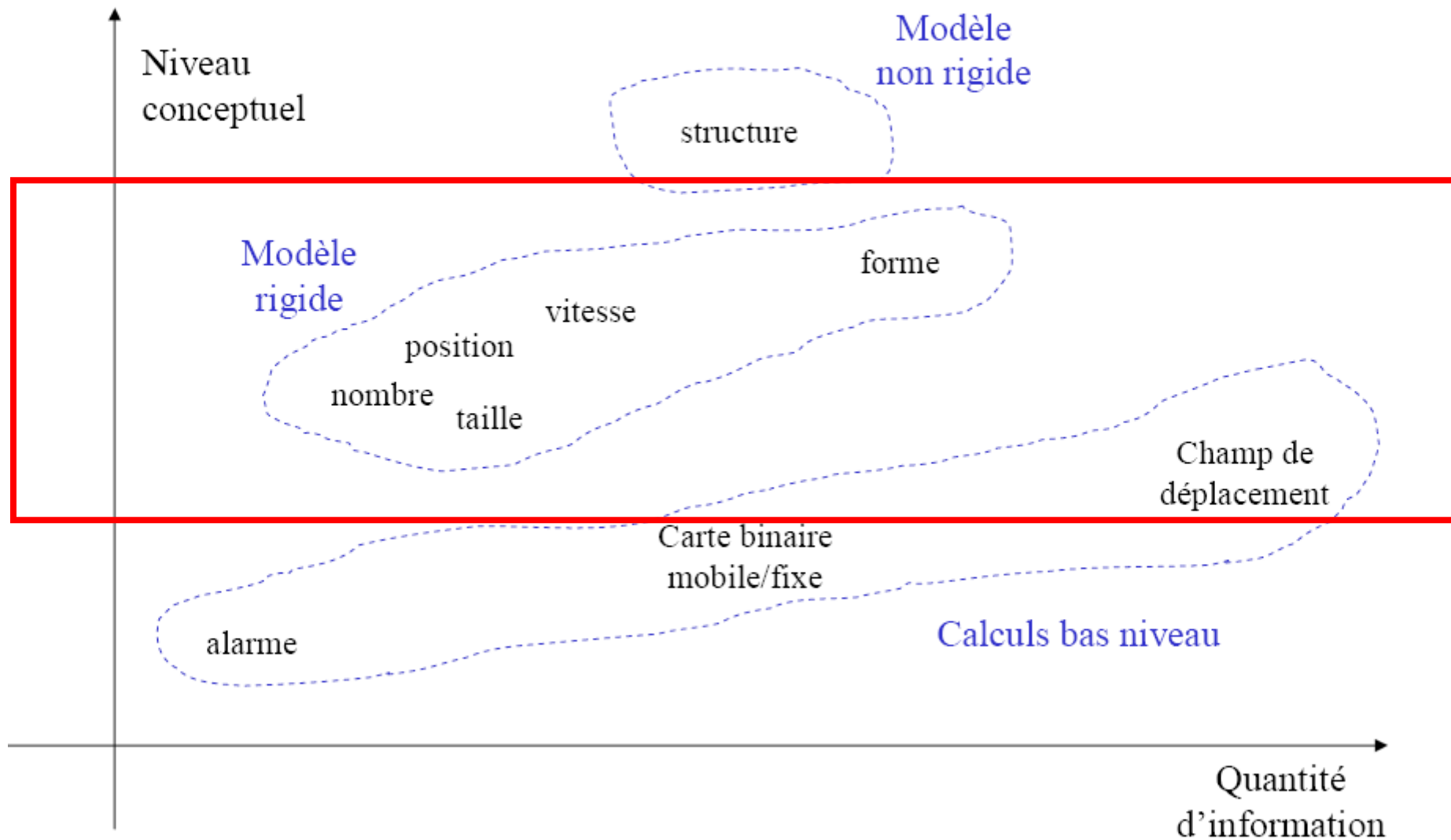
La taille apparente de la barre $l(t)$ à l'instant t sur le plan image s'écrit :

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

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

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

Quelles informations extraire ?



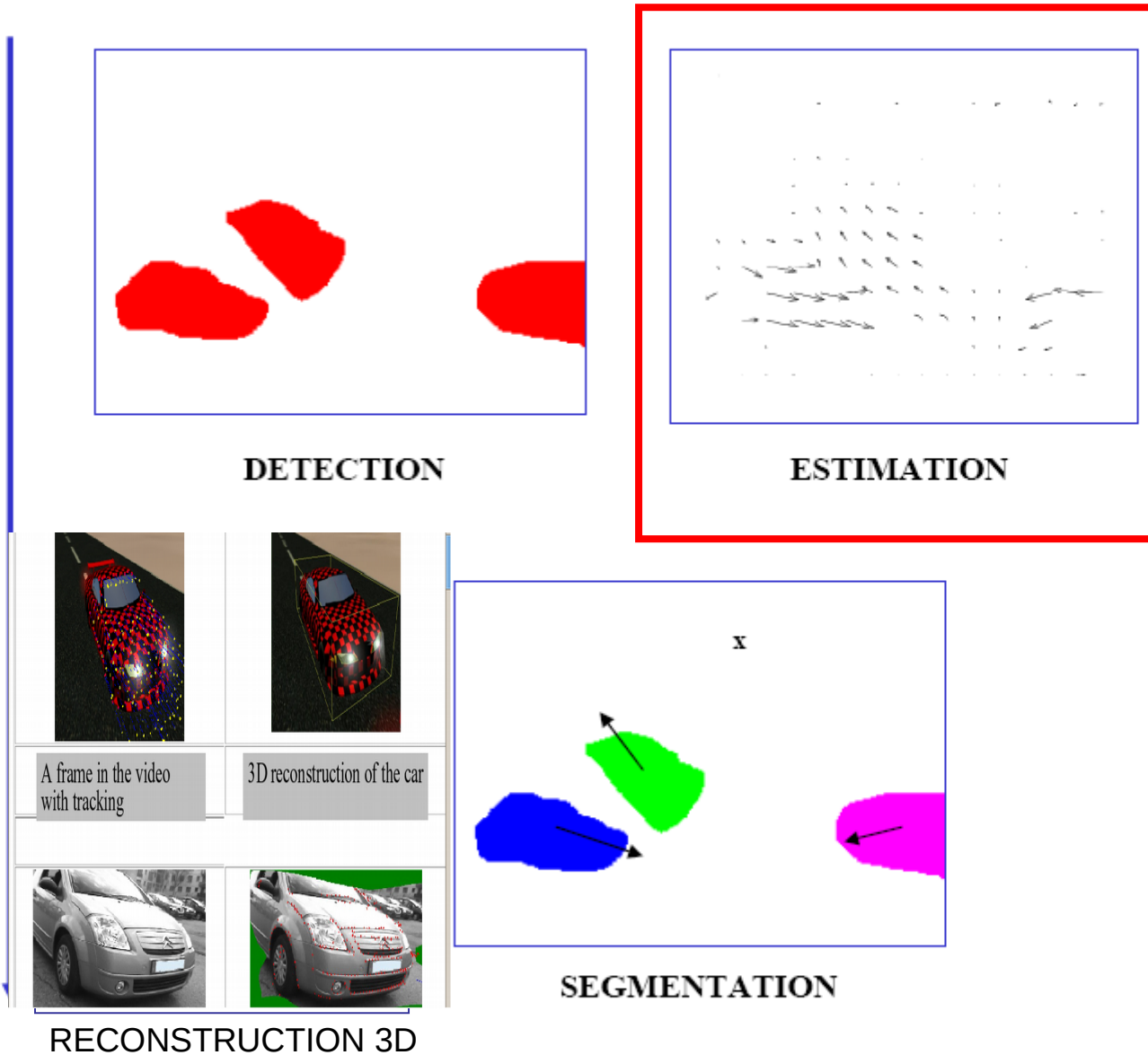
Ce qui se traduit en terme de traitement d'images ...

On distingue en général 3 tâches de traitement d'images :



Taxi_mp4.avi

Niveau conceptuel



Les 2 Problématiques Informatiques

1. Le problème de la mise en correspondances complexe (ou recalage) : https://fr.wikipedia.org/wiki/Recalage_d%27images

Quels éléments d'une trame correspondent à quels éléments de la trame suivante de la séquence ?

<https://fr.wikipedia.org/wiki/RANSAC>
<https://image-matching-workshop.github.io>

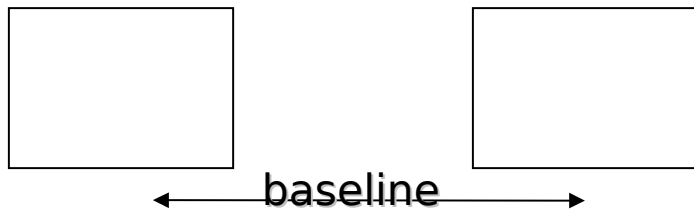
On empruntera deux voies possibles :

- **Les méthodes différentielles** : en sortie, on obtiendra des mesures denses, cad, calculées en chaque pixel image. Elles utilisent des dérivées temporelles du signal et donc nécessitent des séquences finement échantillonnées.
- **Les méthodes de Matching** : en sortie, on obtiendra des mesures éparses, cad, calculées seulement en un sous-ensemble des pixels images -> Filtrage de **Kalman**

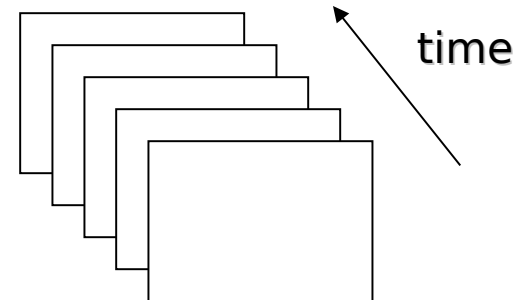
Différence principale entre Mouvement et Stéréo :

Comme les séquences d'image sont généralement échantillonnées temporellement à des taux élevés, les différences spatiales (disparités) entre trames consécutives sont, en moyenne, plus petites que celles de paires stéréo typiques.

- Stereo: Two or more frames

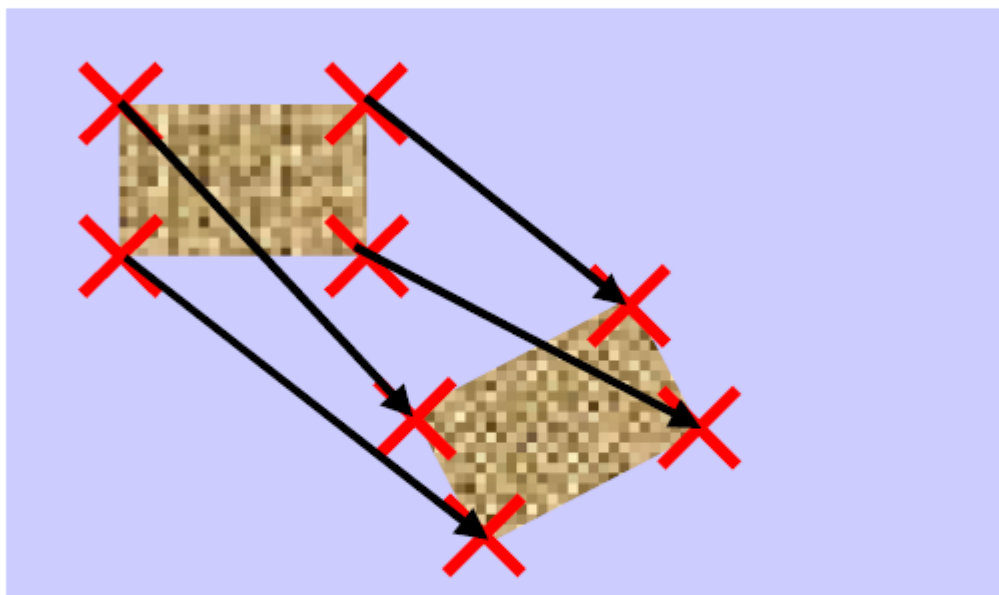


- Motion: N frames



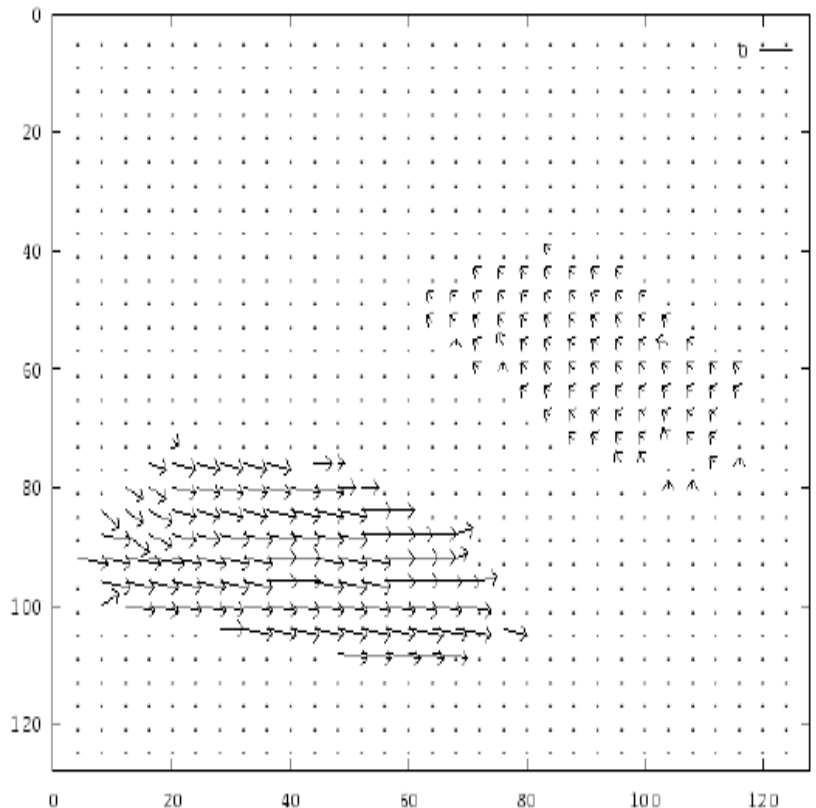
Du coup,

- Grâce aux nombreuses trames échantillonnées de façon proche disponibles pour l'analyse d'image :
 - > les méthodes utilisant des primitives peuvent être plus efficaces grâce aux techniques dites de suivi ou *tracking*, qui exploitent l'historique du mouvement des primitives pour prédire les disparités dans la trame suivante.



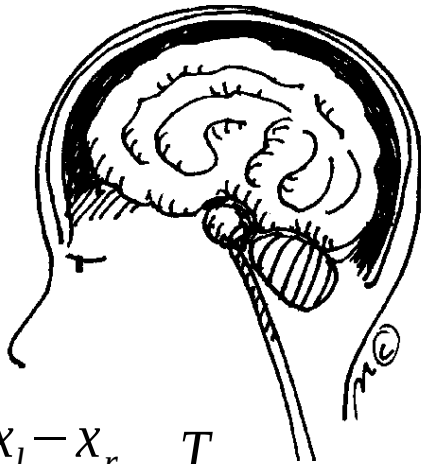
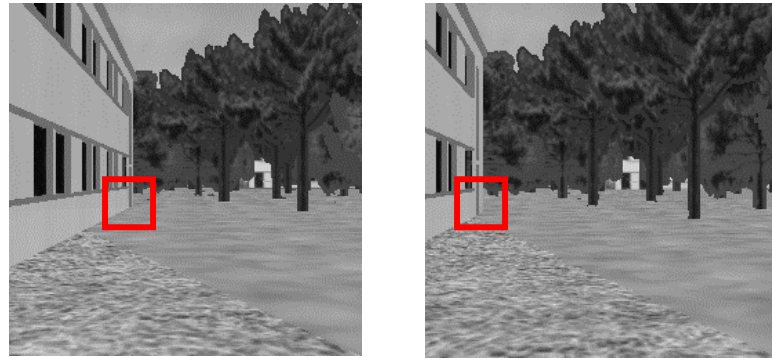
- Grâce aux différences spatio-temporelles généralement petites entre trames consécutives :

-> le problème de correspondance devient le problème de l'estimation du mouvement apparent de motifs lumineux de l'image, que l'on appelle généralement le **flot optique**.



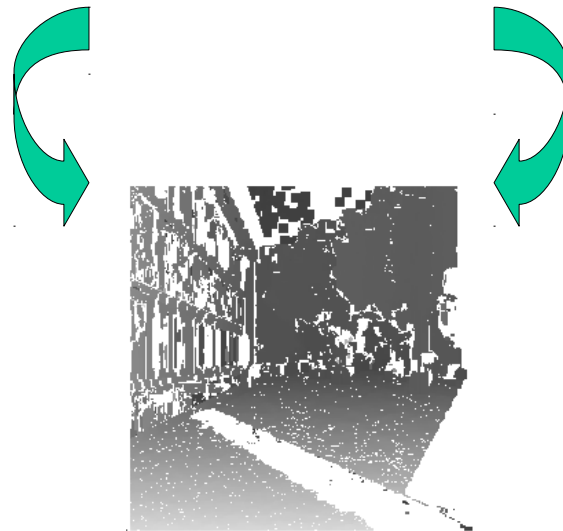
Un parallèle intéressant :

La vision stéréoscopique fait référence à la capacité d'inférer de l'information sur la structure 3D et les distances 3D d'une scène à partir d'au moins deux images optiques prises de points de vue différents



$$\frac{T+x_l-x_r}{Z-f} = \frac{T}{Z}$$

$$d'où \quad Z=f\frac{T}{d} \quad \text{avec} \quad d=x_r-x_l$$



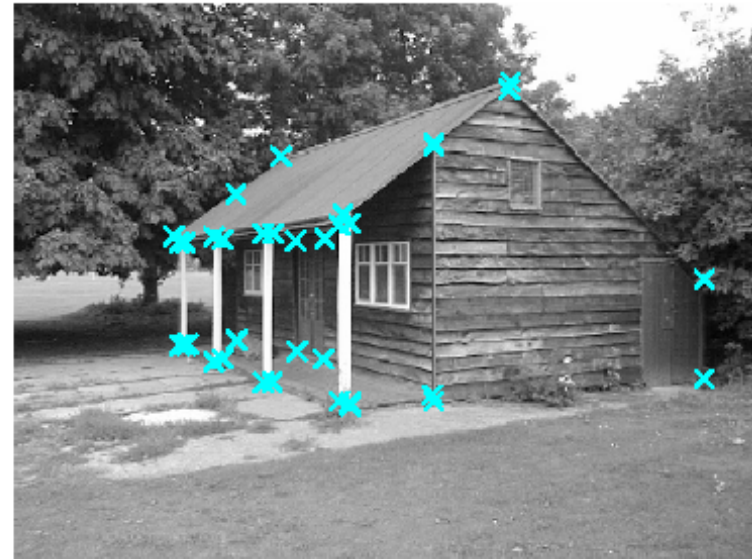
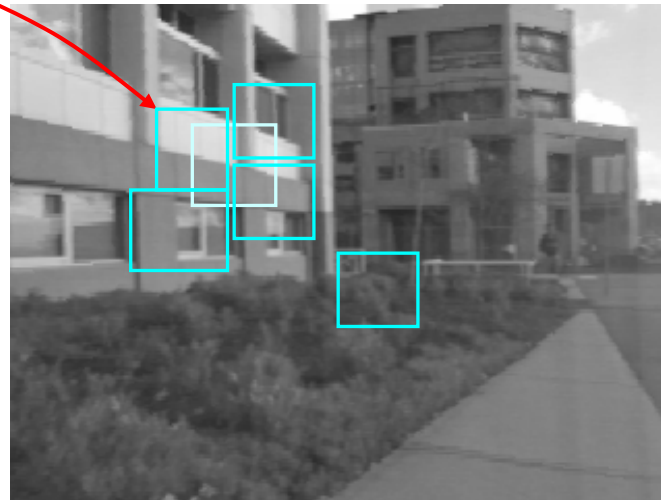
Le problème de la correspondance

Recherche exhaustive ?

Left



Right



Algo CORR_MATCHING

INPUT :

- Paire stéréo d'images I_l et I_r .
- Soient p_l et p_r les pixels dans l'image droite et gauche. $2W+1$ la taille en pixels de la matrice de corrélation. $R(p_l)$ la région de recherche dans l'image droite associée à p_l .
- Soit $\Psi(u,v)$ une fonction de deux valeurs de pixel.

Pour chaque pixel $p_l=[i,j]^T$ de l'image gauche :

- Pour chaque déplacement $d=[d_1,d_2]^T \in R(p_l)$ calculer :

$$c(d) = \sum_{k=-W}^W \sum_{l=-W}^W \Psi(I_l(i+k, j+l), I_r(i+k-d_1, j+l-d_2))$$

- La disparité de p_l est le vecteur $\bar{d}=[\bar{d}_1, \bar{d}_2]^T$ qui maximise $c(d)$ sur $R(p_l)$:

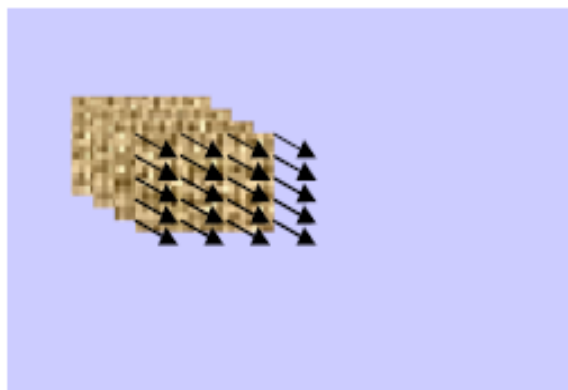
$$\left\{ \mathbf{c}(\mathbf{d}) \right\}$$

OUTPUT : tableau de disparités (disparity map), une pour chaque pixel de I_l .

$$C(x,y,d) = \frac{\sum_{i,j} \left[\left(I_1(x+i, y+j) - \overline{I_1(x,y)} \right) - \left(I_2(x+i+d, y+j) - \overline{I_2(x+d,y)} \right) \right]^2}{\sqrt{\sum_{i,j} \left(I_1(x+i, y+j) - \overline{I_1(x,y)} \right)^2} * \sqrt{\sum_{i,j} \left(I_2(x+i+d, y+j) - \overline{I_2(x+d,y)} \right)^2}}$$

Lien avec la vision biologique :

2 types de perception du mouvement bien distincts chez les primates :



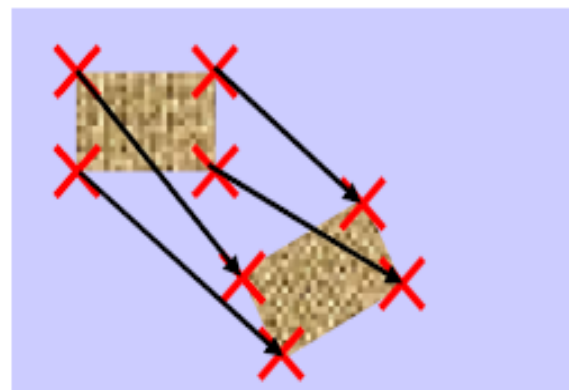
petits mouvements



analyse du flux optique



calcul de champ *dense*



grands mouvements



mise en correspondance de points caractéristiques



calcul de champ *épars*

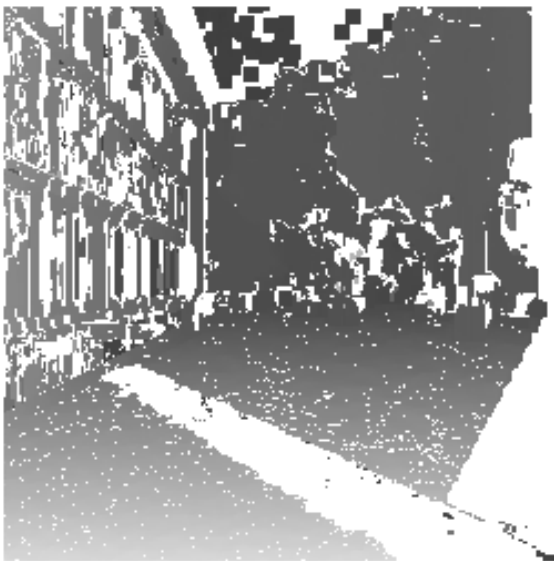
Les 2 Problématiques Informatiques

2. Le problème de la reconstruction :

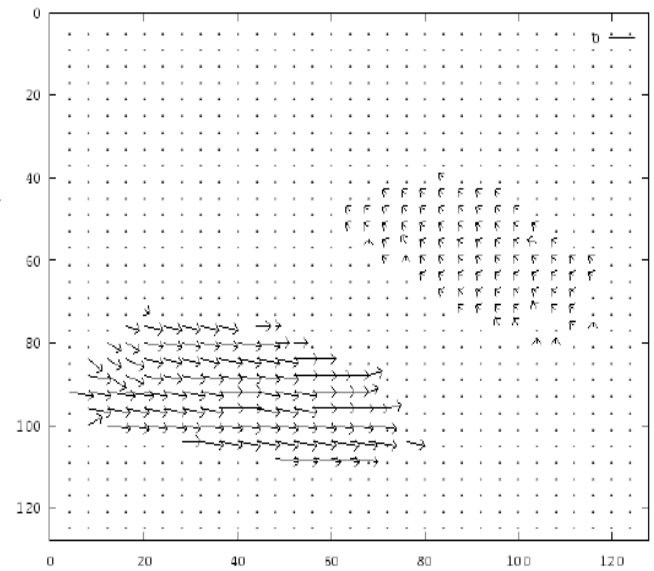
Etant donné un certain nombre d'éléments correspondants, et éventuellement la connaissance des paramètres intrinsèques de la caméra, que pouvons-nous dire au sujet des mouvement et structure 3D de la scène observée ?

Différence principale entre Mouvement et Stéréo :

- Contrairement à la stéréo, dans le mouvement, le déplacement relatif 3D entre la caméra d'observation et la scène observée n'est pas nécessairement causé par une transformation rigide 3D unique.
- La problématique de reconstuction 3D à partir du mouvement est plus délicate que dans le cadre stéréoscopique car il y a une sensibilité accrue au bruit en raison d'une ligne de base entre trames consécutives du point de vue stéréo très petite.



Carte de disparité (à droite) contre flot optique (à droite)



Hypothèse forte de ce cours

Il y a seulement un mouvement rigide relatif entre la caméra et la scène observée, et les conditions d'illuminations ne changent pas dans le temps.

Remarque : ceci implique que les objets 3D observés ne peuvent bouger avec des mouvements différents. Si cela est vraisemblable dans le cas d'un bâtiment observé par un passant se promenant, cette hypothèse est régulièrement violée :

- Matches de foot;
- Traffic routier ou urbain;
- Objets déformables comme les vêtements ou la danse.

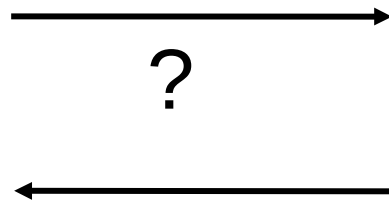
Du coup, si la caméra observe plus d'un objet mobile, ou que simplement vous ne pouvez pas faire l'hypothèse d'une caméra se déplaçant dans un environnement statique, un troisième sous-problème doit être envisagé :

La 3^{ème} problématique informatique

3. Le problème de la **segmentation** :

quelles sont les régions du plan image qui correspondent à différents objets mobiles ?

Matching
Problem



Regions of different moving
objets : Segmentation
Problem

Le champ de déplacement d'objets rigides

Objectifs :

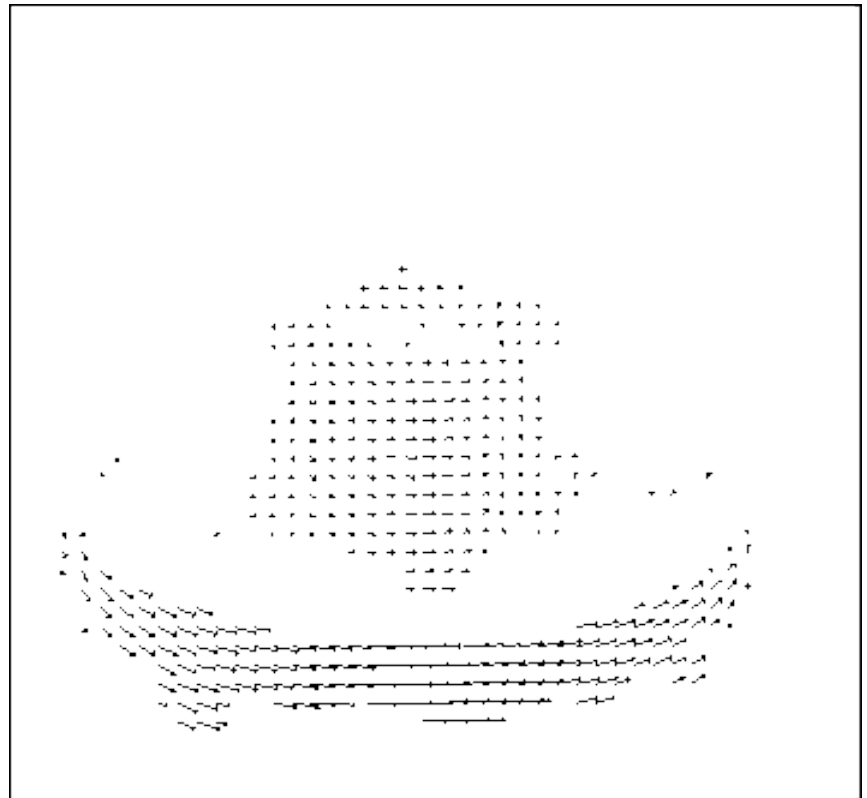
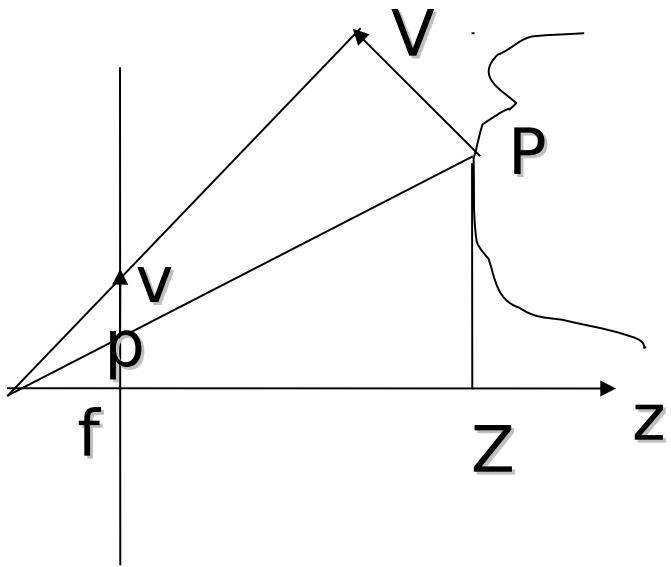
- Prendre contact avec la théorie et la géométrie du **champ de mouvement ou de déplacement** (*motion field*);
- Comparer la représentation des disparités dans le mouvement ou la stéréo;
- Analyser deux cas spéciaux de déplacement rigide;
- [Introduire la notion de **parallaxe de mouvement.**]

Champ de mouvement ou de déplacement

Le champ de mouvement est un champ de vecteurs 2D représentant les vitesses des points images, induit par le mouvement relatif entre la caméra d'observation et la scène observée.

Remarque 1 : ce champ de vitesse peut être vu comme *la projection d'un champ de vitesse 3D sur le plan image*.

Remarque 2 : On travaillera dans cette partie dans le repère caméra, supposant la calibration intrinsèque effectuée.



Notations :

Soit $\mathbf{P}=[X,Y,Z]^T$ un point 3D exprimé dans le référentiel caméra. L'image $\mathbf{p}=[x,y,f]^T=[x,y]^T$ de \mathbf{P} est donnée par :

$$\vec{p} = f \frac{\vec{P}}{Z}$$

Le mouvement relatif entre P et la caméra est décrit par :

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

où \mathbf{T} est la composante en translation du mouvement et $\boldsymbol{\omega}$ la vitesse angulaire de la caméra par rapport à la scène.

Remarque : \mathbf{T} représente un vecteur vitesse, pour une fois, et pas un vecteur déplacement comme d'habitude. De plus, comme le mouvement est rigide, \mathbf{T} et $\boldsymbol{\omega}$ sont les mêmes pour tous les points de l'objet.

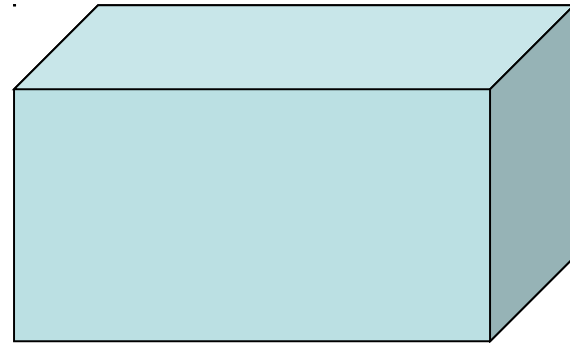
Notations :

L'équation précédente s'écrit aussi en 3D :

$$V_x = -T_x - \omega_y Z + \omega_z Y$$

$$V_y = -T_y - \omega_z X + \omega_x Z$$

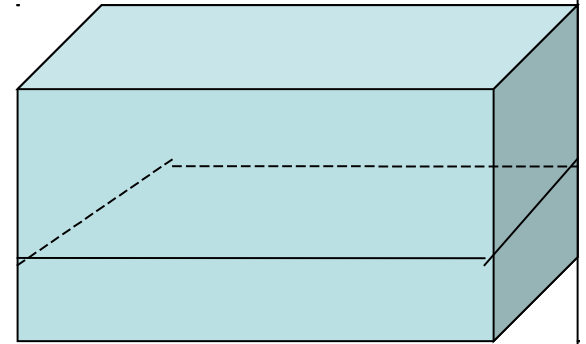
$$V_z = -T_z - \omega_x Y + \omega_y X$$



Les équations de base du Champ de Mouvement

Le champ de mouvement \mathbf{v} est donné par :

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



Soit en 2D :

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

Ce champ de mouvement est la somme de deux composantes, l'une uniquement translationnelle v^T et l'autre uniquement rotationnelle v^ω :

$$\vec{v} = \vec{v}^T + \vec{v}^\omega$$

Soit :

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

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

et

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

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

D'où :

- $v = [v_x, v_y, 0]^T = [v_x, v_y]^T$

- La partie du champ de mouvement qui dépend de la vitesse angulaire ne porte pas d'information de profondeur : les valeurs Z et ω sont découplées.

Remarque : si la mise en correspondance reste un enjeu majeur en stéréo et en mouvement, les déplacements de points sont représentés par des **cartes de disparité** pour la stéréo et par des **champs de mouvement** plus complexe pour le mouvement.

Carte de disparité stéréo vs. Champ de mouvement

Les déplacements spatiaux de points correspondants entre images d'une paire stéréo (formant la carte de disparité stéréo) sont finis, et, en principe, non contraints.

Les déplacements spatiaux de points correspondants entre trames consécutives d'une séquence dynamique (formant le champ de mouvement) sont des approximations discrètes de dérivées temporelles, et doivent donc être relativement faibles.

Le champ de mouvement coïncide avec la carte de disparité stéréo seulement si les différences spatiales et temporelles entre trames consécutives sont suffisamment faibles.

Le cas particulier de la Translation Pure $V=-T$.

Le champ de mouvement correspondant possède une structure spatiale particulière.

Comme $\omega=0$:

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

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

On considère le cas particulier où $T_z \neq 0$. On introduit alors le point $p_0 = [x_0, y_0]^T$ défini par :

$$x_0 = \frac{f T_x}{T_z}$$

$$\text{alors } v_x = (x - x_0) \frac{T_z}{Z}$$

$$y_0 = \frac{f T_y}{T_z}$$

$$v_y = (y - y_0) \frac{T_z}{Z}$$

Le cas particulier de la Translation Pure $V=-T$.

Le champ est donc **radial** : les vecteurs irradiant d'une origine commune p_0 qui est donc le **point de fuite** de la direction de translation.

Donc,

- si $T_z < 0$, $V_z > 0$ et le point s'éloigne de la caméra. Le champ de vecteurs pointent vers p_0 qui est appelé le **focus de contraction**.
- si $T_z > 0$, $V_z < 0$ et le point se rapproche de la caméra. Le champ de vecteurs pointent à partir de p_0 qui est appelé le **focus d'expansion**.

$$\|v(p)\| = \frac{\|T_z\|}{Z} \|p - p_0\|$$

• et

Le cas particulier de la Translation Pure $V=-T$.

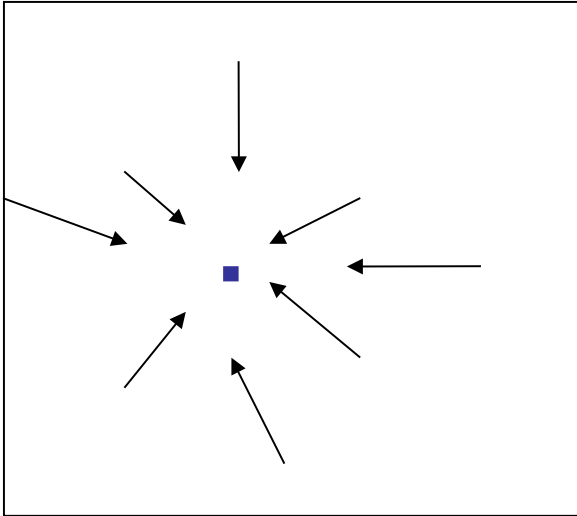
$$\begin{cases} v_x = -f \frac{T_x}{Z} \\ v_y = -f \frac{T_y}{Z} \end{cases}$$

• si $T_z=0$,

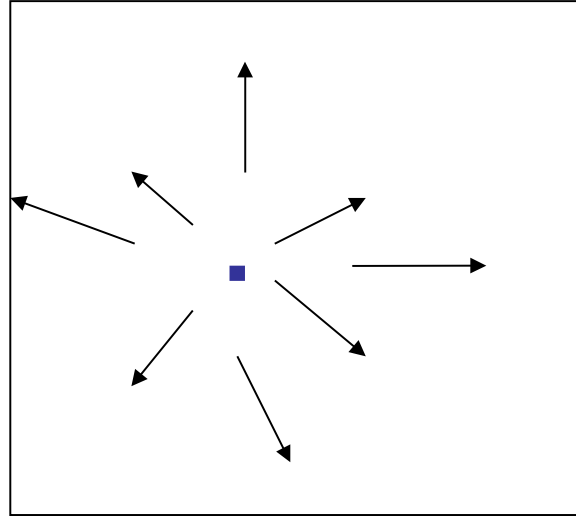
et tous les vecteurs du champ de mouvement sont parallèles
avec

$$\|v(p)\| = f \frac{\|T\|}{Z} = \text{Constante}$$

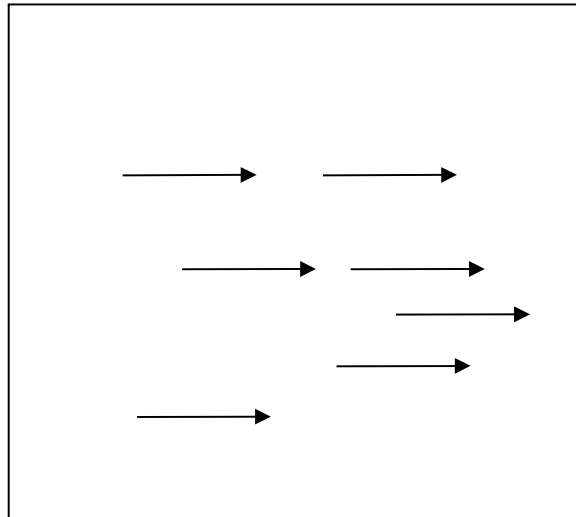
L'objet et la caméra s'éloignent



L'objet et la caméra se rapprochent



L'objet et la caméra restent à distance constante



Le cas particulier de la Translation Pure $V=-T$.

- Si $T_z \neq 0$, le champ de mouvement est **radial**, et tous les vecteurs pointent vers (ou à partir de) un seul point, p_0 . Si $T_z = 0$, le champ de mouvement est **parallèle**.
- La longueur des vecteurs du champ de mouvement est inversement proportionnel à la profondeur Z ; Si $T_z \neq 0$, elle est aussi proportionnel à la distance entre p et p_0 ;
 - p_0 est le point de fuite de la direction de translation.
 - p_0 est aussi le point d'intersection du rayon optique parallèle à la direction de translation avec le plan image.

Le cas particulier du plan mobile

Penser à un environnement construit par l'homme.
On suppose que la caméra observe une surface plane, π ,
d'équation :

$$\vec{n}^T \vec{P} = d$$

où $\mathbf{n}=[n_x, n_y, n_z]^T$ est le vecteur unité normal à π , et d la distance entre l'origine et π .

Le plan se déplace avec une vitesse translationnelle $-T$ et une vitesse angulaire $-\omega$, si bien que \mathbf{n} et d dépendent du temps. On a :

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

Le cas particulier du plan mobile

En remplaçant Z dans les équations de base du champ de mouvement :

$$\begin{cases} v_x = \frac{1}{fd} (a_1 x^2 + a_2 xy + a_3 fx + a_4 fy + a_5 f^2) \\ v_y = \frac{1}{fd} (a_1 xy + a_2 y^2 + a_6 fy + a_7 fx + a_8 f^2) \end{cases}$$

$$a_1 = -d\omega_y + T_z n_x$$

$$a_2 = d\omega_x + T_z n_y$$

$$a_3 = T_z n_z - T_x n_x$$

$$a_4 = d\omega_z - T_x n_y$$

$$a_5 = -d\omega_y - T_x n_z$$

$$a_6 = T_z n_z - T_y n_y$$

$$a_7 = -d\omega_z - T_y n_x$$

$$a_8 = d\omega_x - T_y n_z$$

Le champ de mouvement d'une surface plane mobile, à tout instant t , est un polynôme quadratique en les coordonnées (x,y,f) des points images.

Symétrie des coefficients $\rightarrow a_i$ inchangés si

$$d' = d$$

$$n' = T \parallel T \parallel$$

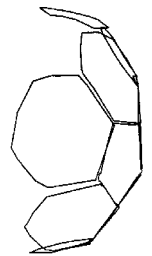
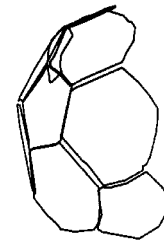
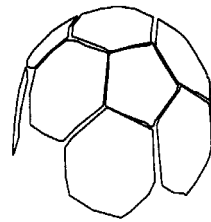
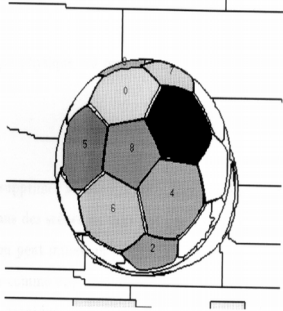
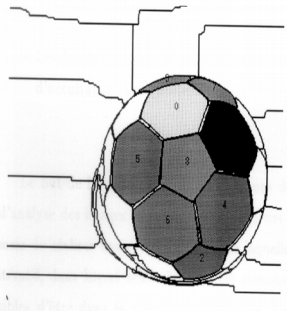
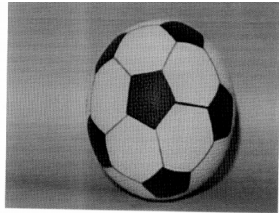
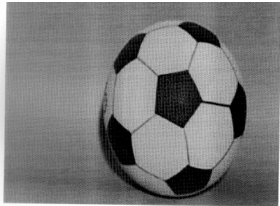
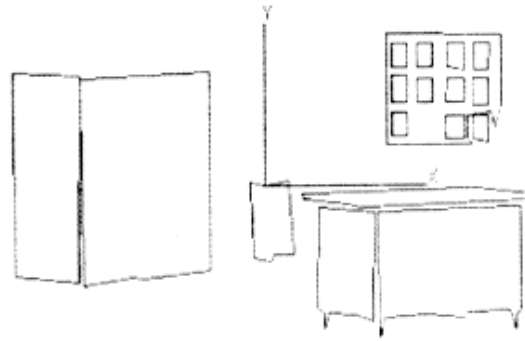
$$T' = \parallel T \parallel n$$

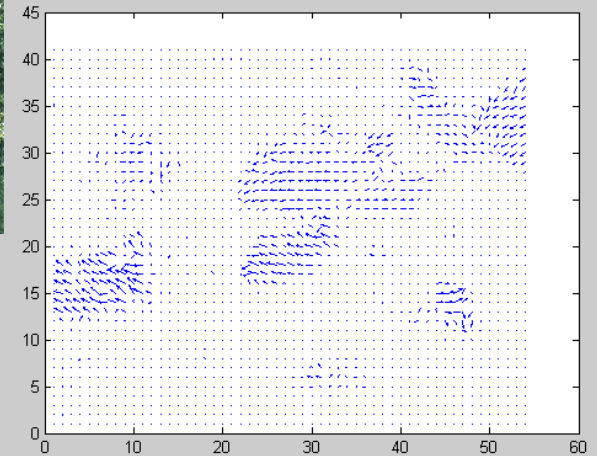
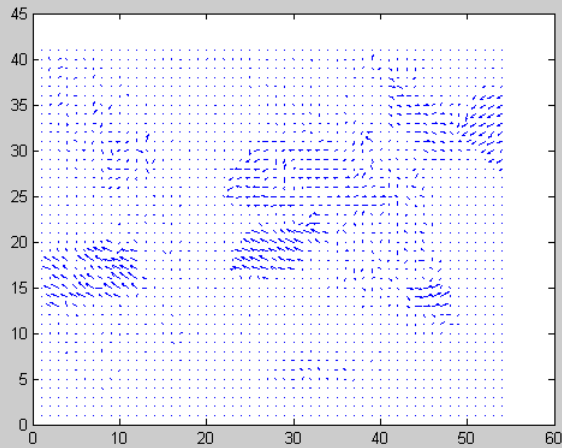
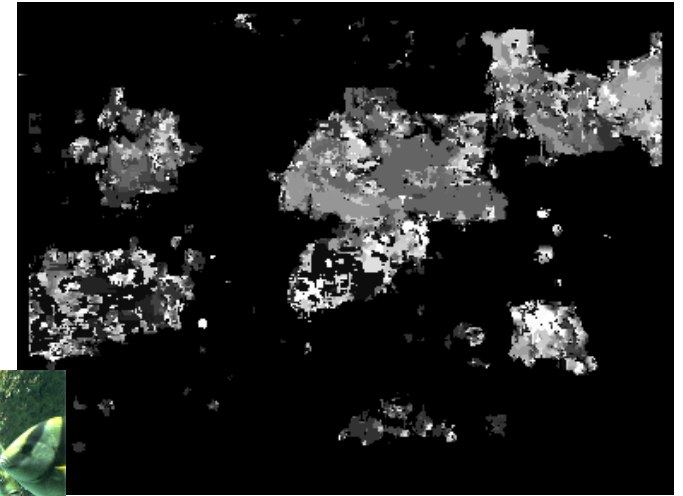
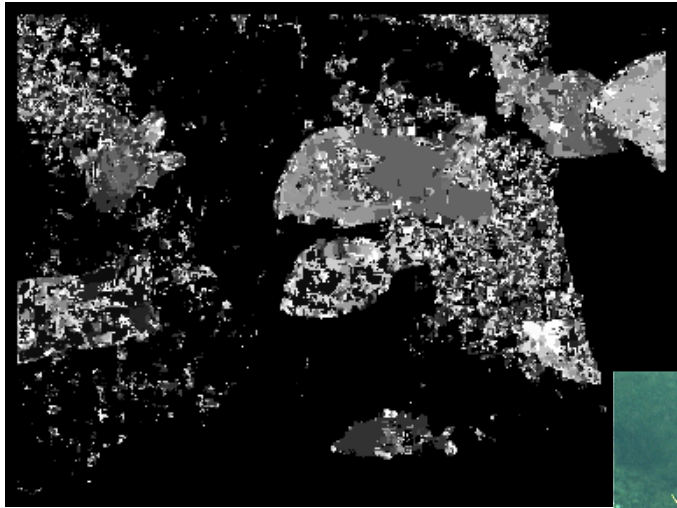
$$\omega' = \omega + n \times T / d$$

En général, le même champ de mouvement peut être produit par deux plans différents avec deux mouvements 3D différents ! \rightarrow Indétermination pour le problème inverse !

2 conclusions importantes :

- Puisque le champ de mouvement d'une surface plane est décrite exactement et globalement par un polynôme du second degré, le champ de mouvement de toute surface lisse est approximable par un polynôme de bas ordre même sur de larges régions du plan image. Et donc, en général, des modèles paramétriques très simples permettent une estimation assez précise du champ de mouvement.
- Comme les algos recouvrant le mouvement et la structure 3D ne peuvent utiliser des points coplanaires (indétermination), les mesures devront être effectuées en de nombreux différents emplacement du plan image pour minimiser la proba d'examiner des points se trouvant proche d'une surface plane.





Représentation des normes des vecteurs vitesses et du flot optique évalués par appariement à partir d'images en niveaux de gris à gauche, à partir d'images en couleurs à droite

Le plan mobile : propriétés du champ de mouvement

1. Le champ de mouvement d'une surface plane est, à tout instant, un polynôme quadratique en les coordonnées images.
2. En raison de la symétrie particulière des coefficients polynomiaux, le même champ de mouvement peut être produit par deux surfaces planes différentes avec deux mouvements 3D distincts.

On a vu comment, on pouvait déduire la forme du champ de déplacement 2D si on connaît le champ de vitesses 3D et les paramètres caméra.

On suppose donc que la connaissance de champ 2D nous donnera des indications sur les mouvements réels 3D et sur les structures 3D présentes dans la scène.

La question qui vient naturellement est : comment estimer ce champ déplacements 2D ?

La notion de flot optique apportera un élément de réponse.

La notion de Flot Optique

Nous allons bientôt essayer de voir comment estimer *le champ de mouvement à partir d'une séquence d'images*

Pour nous une séquence d'images signifie un ensemble de variations spatio-temporelles de la luminosité d'une image.

Donc, il faut trouver le lien physique entre les variations lumineuses et le champ de mouvement.

1ère Hypothèse de cette partie : de petits mouvements

La luminosité image est continue et différentiable autant de fois que nécessaire à la fois dans l'espace et dans le temps.

Calcul du flot optique : limites et contraintes (brightness constancy)

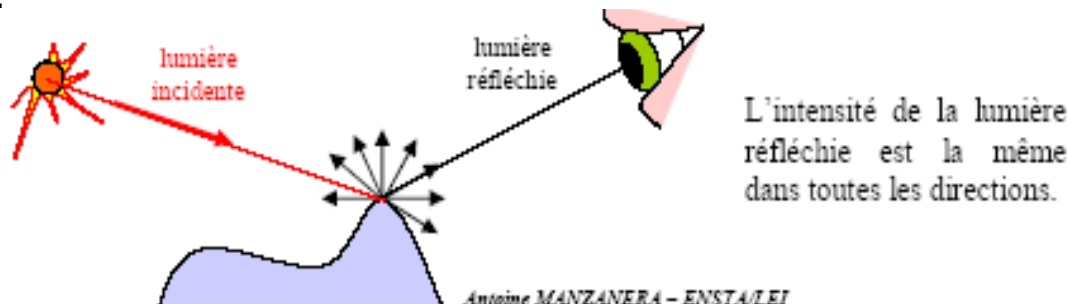


(1) On suppose :

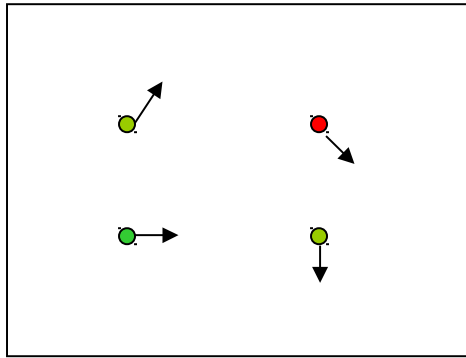
MOUVEMENT \Leftrightarrow VARIATION D'INTENSITE

(\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'éclaircement constant sur la scène.

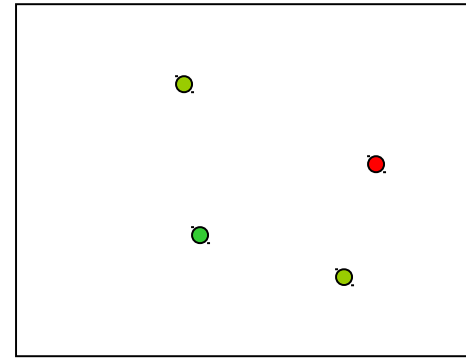
(\Rightarrow) On suppose qu'un mouvement induit un déplacement d'intensité dans l'image. Pour pouvoir utiliser la variation d'intensité dans l'image pour retrouver le mouvement apparent de l'objet dans la scène, on doit donc supposer que la lumière réfléchie par un point de la scène reste constante indépendamment du mouvement relatif scène/caméra. L'hypothèse sous-jacente est que les objets sont à surface lambertienne :



Donc, un mouvement induit un déplacement spatial d'une région d'intensité lumineuse dans l'image sans modification d'intensité dans cette région.



$H(x, y)$



$E(x, y)$

Comment estimer le mouvement d'un pixel entre l'image H et l'image E ?

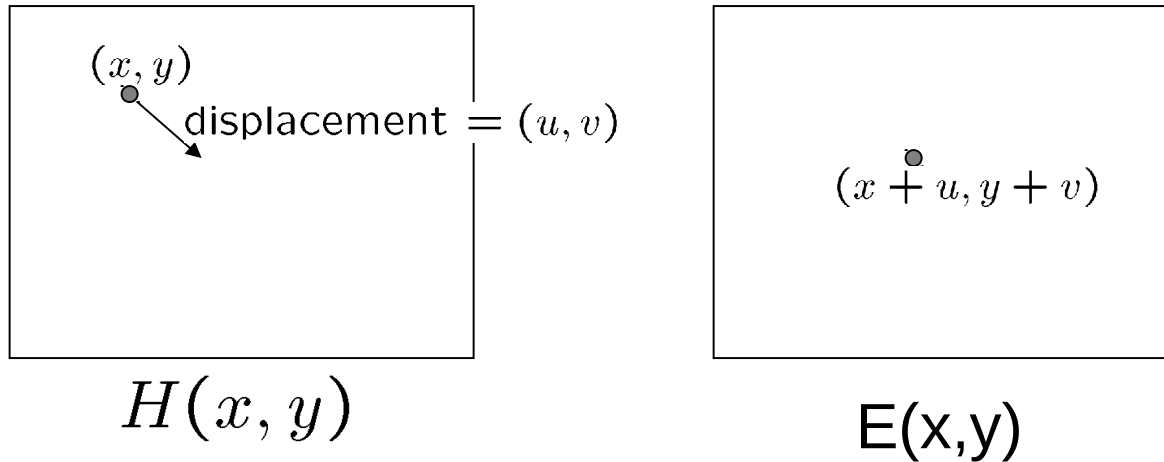
Donc si $\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t} \Leftrightarrow \frac{\Delta I(x, y)}{\Delta t}$,

selon les hypothèses précédentes, cela revient résoudre le problème de la m.e.c. pixellique :

- Étant donné un pixel dans l'image H, chercher des pixels voisins de la même couleur dans l'image E

C'est ce qu'on appelle le problème du **flot optique**

Calcul du flot optique : limites et contraintes (brightness constancy)



- Illumination constante : $0 = E(x + u, y + v) - H(x, y)$
- Des petits mouvements : $u = \Delta x$ et $v = \Delta y$ valent moins que 1 pixel : prenons alors le développement en série de Taylor d'ordre 1 de E :

$$E(x + \Delta x, y + \Delta y) = E(x, y) + \frac{\partial E}{\partial x} \Delta x + \frac{\partial E}{\partial y} \Delta y + \text{termes d'ordres supérieurs}$$

$$E(x + \Delta x, y + \Delta y) \approx E(x, y) + \frac{\partial E}{\partial x} \Delta x + \frac{\partial E}{\partial y} \Delta y$$

En combinant les deux équations :

$$0 = E(x + \Delta x, y + \Delta y) - H(x, y)$$

$$0 \approx E(x, y) + E_x \Delta x + E_y \Delta y - H(x, y) \text{ avec } E_x = \frac{\partial E}{\partial x}$$

$$0 \approx (E(x, y) - H(x, y)) + E_x \Delta x + E_y \Delta y$$

$$0 \approx \Delta E + E_x \Delta x + E_y \Delta y$$

$$0 \approx \Delta E + \vec{\nabla} E \cdot [\Delta x \quad \Delta y]$$

$$0 \approx \frac{\Delta E}{\Delta t} + \vec{\nabla} E \cdot \left[\frac{\Delta x}{\Delta t} \quad \frac{\Delta y}{\Delta t} \right]$$

A la limite, quand Δx et Δy tendent vers 0, cela s'écrit différentiellement :

$$E_t + \vec{\nabla} E \cdot \left[\frac{dx}{dt} \quad \frac{dy}{dt} \right] = 0$$

Cela se traduit par l'équation fondamentale suivante décrivant la luminosité image E :

$$\frac{dE(x, y, t)}{dt} = 0$$

x et y variant dans le temps :

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

C'est la traduction sous forme d'équation de l'expérience commune selon laquelle dans la plupart des circonstances, la luminosité apparente d'objets mobiles reste constante.

Soit

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial x} \\ \frac{\partial E}{\partial y} \end{bmatrix}$$

(Frame spatial gradient)

$$v = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}$$

(optical flow)

et

$$E_t = \frac{\partial E}{\partial t}$$

(derivative across frames)

Rappels :

Dérivée du premier ordre d'une fonction de deux variables $f(x,y)$

= Vecteur Gradient en $P(x,y)$

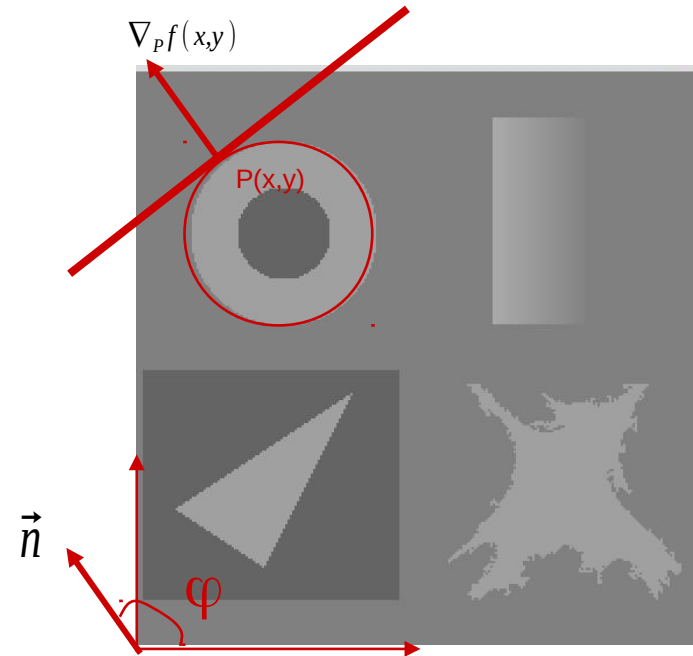
$$\nabla_P f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

= Vecteur normal à la courbe de niveau
 $f(x,y) = f(x_p, y_p) = \text{cste}$, \vec{n}

Orientation du gradient :

$$\phi = \text{Arc tan} \left(\frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial x}} \right)$$

Orientation du contour : $\theta = \frac{\pi}{2} + \phi$



L'équation de Conservation de l'Illumination - *Brighthness Constancy*

Etant donnée la luminosité image $\mathbf{E}(\mathbf{x},\mathbf{y},\mathbf{t})$ et le champ de mouvement \mathbf{v} , on a :

$$\left(\nabla E \right)^T \mathbf{v} + E_t = 0$$

où E_t désigne la dérivée partielle par rapport au temps.

Sous l'hypothèse de petits mouvements, ∇E et E_t sont accessibles par la mesure, donc \mathbf{v} n'est pas loin. Qu'en est-il précisément ?

Comment estimer le champ de mouvement \mathbf{v} à partir de cette équation ?

Peut-on estimer le champ de mouvement **v total** à partir de cette équation ?

Le problème de l'ouverture

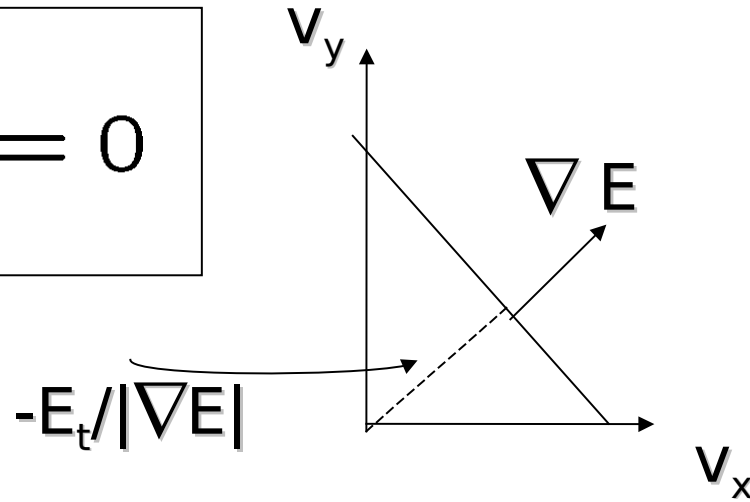
La composante du champ de mouvement dans la direction orthogonale au gradient spatial de l'image n'est pas contrainte par l'équation de conservation de l'illumination.

Donc, seule la composante normale v_n de v dans la direction du gradient spatial peut être estimée à partir de cette équation :

$$\frac{(\nabla E)^T v}{\|\nabla E\|} = -\frac{E_t}{\|\nabla E\|} = v_n$$

Géométriquement,

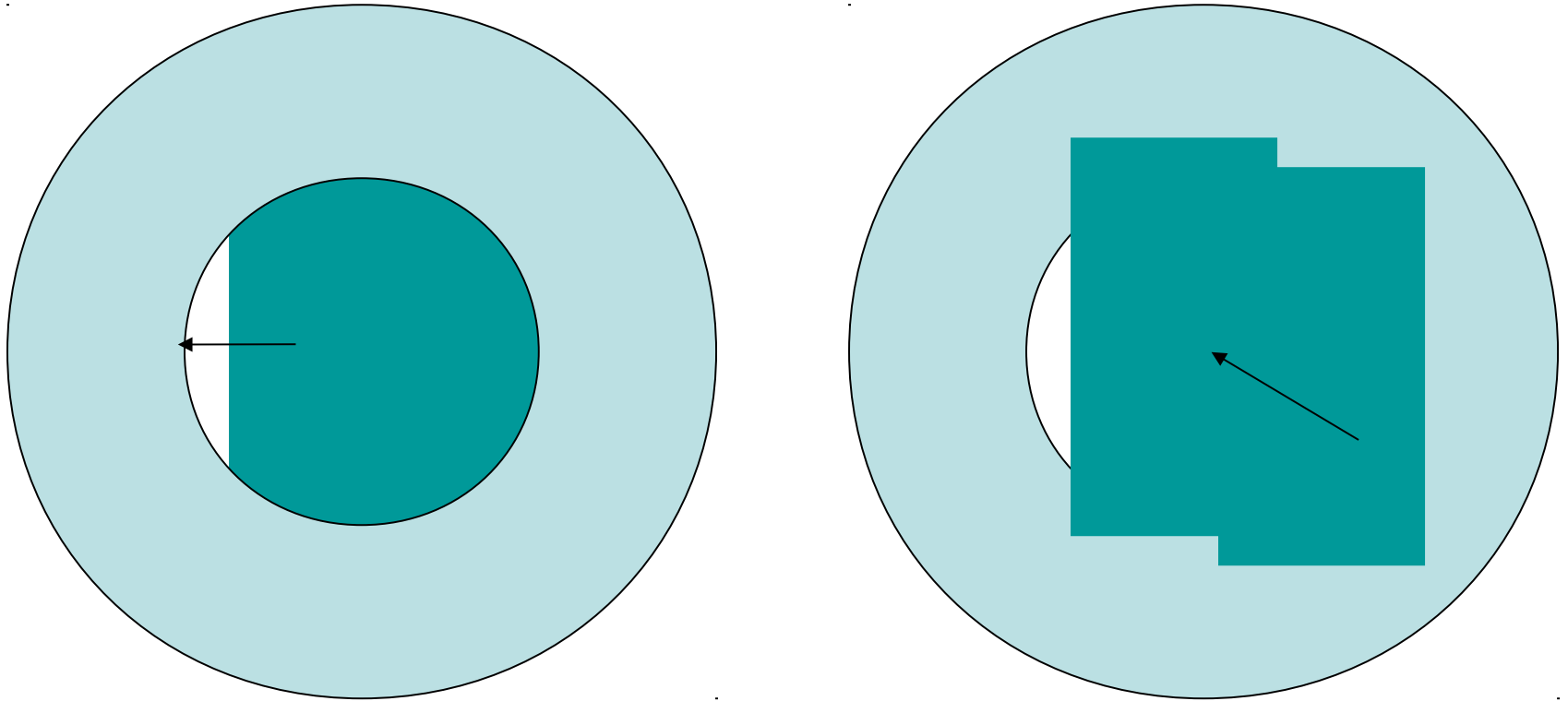
$$(\nabla E)^T \cdot v + E_t = 0$$



Pour pouvoir mesurer ∇E , cela implique qu'on observe de petits mouvements et donc qu'on porte une contrainte sur v dans une petite fenêtre d'analyse.

$$\frac{\partial E}{\partial x} \rightarrow \frac{\Delta E}{\Delta x}$$

Le problème de l'ouverture



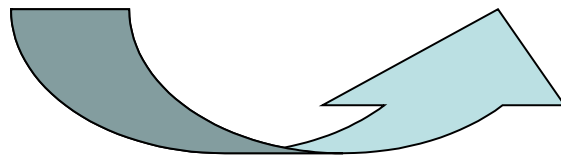
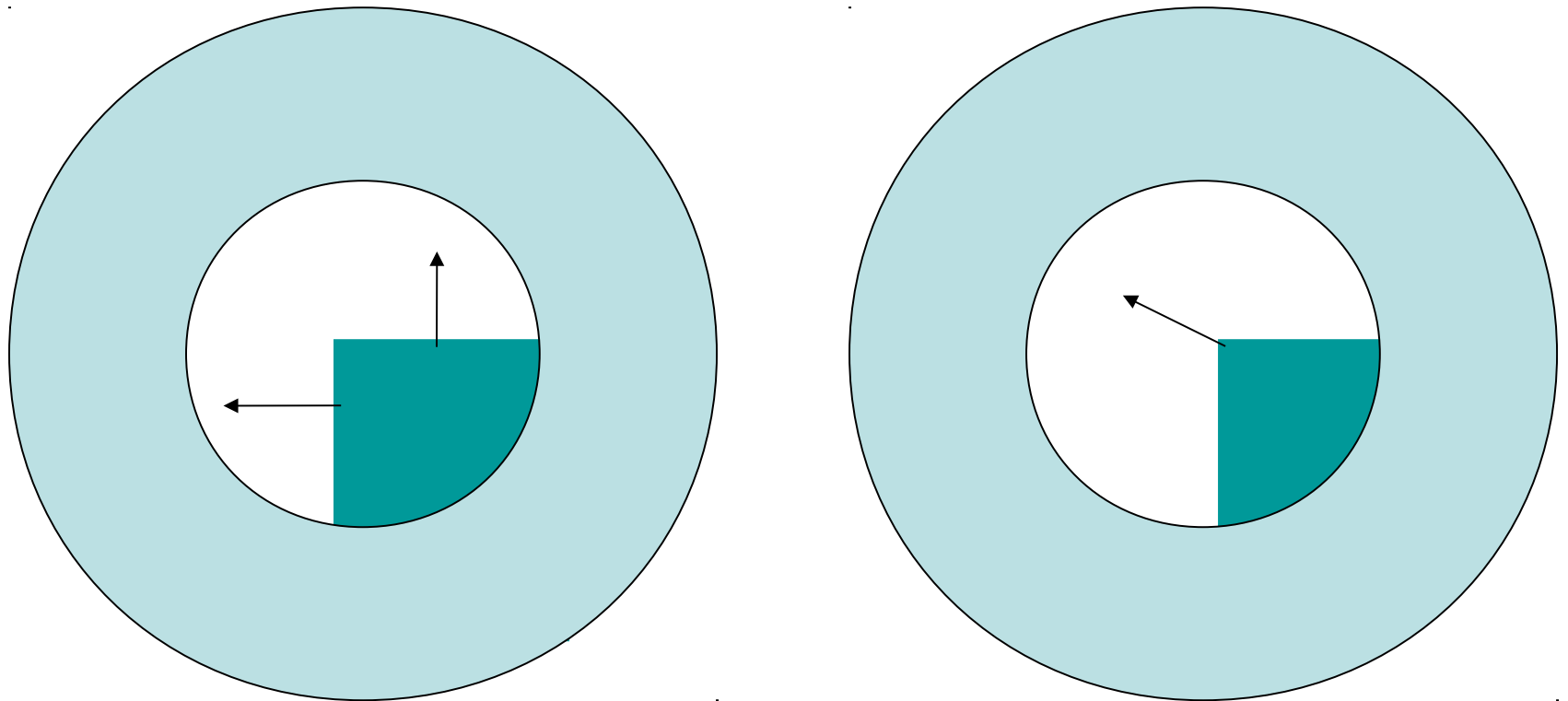
The Image Brightness Constancy Assumption only provides the OF component in the direction of the spatial image gradient

Le parallèle entre l'illustration visuelle précédente et l'équation de conservation de l'illumination n'est pas parfait.

En effet, cette équation relie le gradient image et le champ de mouvement en un même point image, établissant une contrainte sur un support spatial infiniment petit, tandis que la figure précédente fait le lien sur un support petit mais fini.

Donc, une stratégie possible pour résoudre le problème d'ouverture est d'examiner les variations spatio-temporelles de l'illumination sur un voisinage de chaque point

-> Solution manifestement adoptée par le système visuel humain.



Moyenne

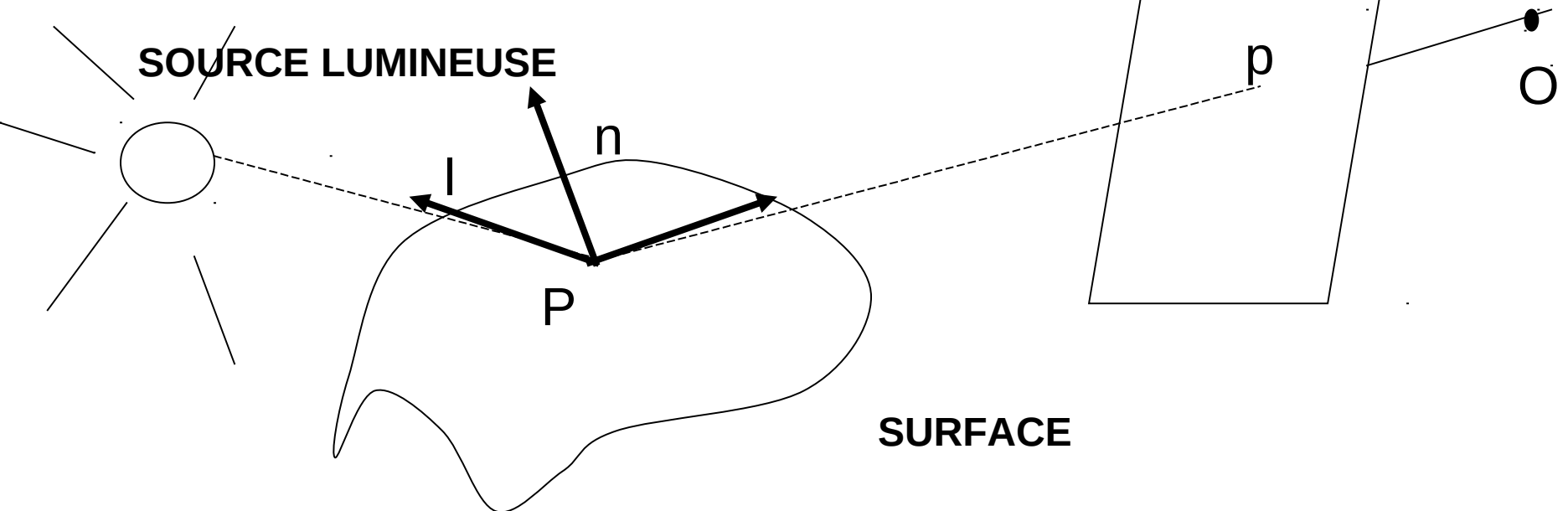
Un autre problème se pose ...

Validité de l'équation de conservation : flot optique

A quel point l'estimée de v_n à partir de l'équation de conservation est-elle correcte ?

Si on se restreint à une surface lambertienne illuminée par une source lumineuse ponctuelle située à l'infini, on peut écrire l'illumination image sous la forme :

$$E = \rho \vec{I}^T \cdot \vec{n}$$



Le problème est que la vitesse rotationnelle d'un objet 3D vient contredire l'hypothèse d'illumination constante (un déplacement d'un point dans l'image ne doit pas modifier sa luminosité) :

$$\text{En effet, } \frac{d(\rho \vec{I}^T \cdot \vec{n})}{dt} = \rho \vec{I}^T \cdot \frac{d\vec{n}}{dt} = \rho \vec{I}^T \cdot (\vec{\omega} \times \vec{n})$$

$$\text{Donc } (\nabla E)^T v + E_t = \rho \vec{I}^T \cdot (\vec{\omega} \times \vec{n})$$

Par conséquent, la différence Δv entre la vraie valeur de v_n après modélisation radiométrique et celle estimée grâce à l'équation de conservation v_n peut s'écrire :

$$|\Delta v| = \rho \frac{|\vec{I}^T \cdot (\vec{\omega} \times \vec{n})|}{\|\nabla E\|}$$

Validité de l'équation de conservation : flot optique

En conséquence, l'équation de conservation de la luminosité fournit la véritable valeur de la composante normale v_n du champ de mouvement ($\Delta v = 0$ pour tout type de surface) seulement pour :

- Des mouvements purement translationnels;
- Tout mouvement rigide pour lequel \mathbf{l} et $\boldsymbol{\omega}$ sont parallèles.

De plus $\Delta v \uparrow$ pour $\|\nabla E\| \downarrow$

Enfin, Δv est rarement nul, et le mouvement apparent de la luminosité image est presque toujours différent du champ de mouvement. Pour cette raison, et pour éviter toute confusion, on appelle le mouvement apparent un flot optique, et on fait référence aux techniques d'estimation du champ de mouvement à partir de l'équation de conservation de l'illumination comme les techniques de flot optique.

Flot Optique

Le flot optique est un champ de mouvement satisfaisant la contrainte de conservation de l'illumination :

$$(\nabla E)^T \cdot v + E_t = 0$$

et pouvant être défini comme le mouvement apparent des motifs lumineux image.

Remarque 1 : le flot optique est différent du champ de déplacement projeté 3D-> 2D mais en est une approximation

Remarque 2 : si les mouvements sont suffisamment petits, on peut estimer la composante normale v_n de ce flot optique.

Flot Optique et Champ de Mouvement

Le *flot optique* est l'*approximation du champ de mouvement* calculée à partir d'une séquence d'images temporelle. Sous les hypothèses simplificatrices suivantes :

- surfaces lambertiennes,
- source lumineuse ponctuelle à l'infini,

l'erreur liée à cette approximation est

- Faible en les points présentant un fort gradient spatial
- Nulle seulement pour les mouvements de translation ou les mouvements rigides telles que la direction de l'illumination est parallèle à la vitesse angulaire.

Estimation du champ de Mouvement

Nous allons maintenant essayer de voir comment estimer *le champ de mouvement à partir d'une séquence d'images*

Les techniques algorithmiques disponibles peuvent être grossièrement divisés en deux catégories :

- Techniques de mises en correspondance dense dites **techniques différentielles** :
-> **méthodes de calcul du flot optique**

Vrai flot optique

- Techniques de mises en correspondances éparses dites **techniques de matching** :
-> **méthodes de tracking**

Méthodes qui peuvent s'affranchir de l'hypothèse de "brightness constancy" car on utilisera des contraintes géométriques

Techniques Différentielles : calcul du Flot Optique Direct par Moindre Carrés

Nous allons exposer une première méthode utilisant l'estimation locale aux Moindres Carrés des paramètres du flot optique

Hypothèses et contraintes

- L'équation de conservation de la luminosité image fournit une bonne approximation de la composante normale du champ de mouvement;
- Le champ de mouvement est bien approximé par un vecteur mouvement *constant* à l'intérieur d'un tout petit patch du plan image. (cf. approximation de champ de mouvement lisse pour des plans mobiles)

$$\begin{cases} v_x = \frac{1}{fd} (a_1 x^2 + a_2 xy + a_3 fx + a_4 fy + a_5 f^2) \\ v_y = \frac{1}{fd} (a_1 xy + a_2 y^2 + a_6 fy + a_7 fx + a_8 f^2) \end{cases}$$

Donc à l'ordre 1, (v_x, v_y) peut être assimilé à un vecteur constant $(a_5 f/d, a_8 f/d)$ localement qui dépend du mouvement générique d'un plan mobile

Pour tout point p_i à l'intérieur d'un petit patch Q de taille $N \times N$ (disons 5×5) on peut écrire :

$$\left(\nabla E(p_i) \right)^T v(p_i) + E_t(p_i) = 0$$

$$\text{et } \forall i, v(p_i) = v_Q$$

$$\text{d'où } \forall p_i \in Q, \left(\nabla E(p_i) \right)^T v_Q + E_t(p_i) = 0$$

En conséquence, le flot optique peut être estimé sur Q comme le vecteur constant v' minimisant la fonctionnelle :

$$\Psi(v) = \sum_{p_i \in Q} \left[\left(\nabla E(p_i) \right)^T v + E_t(p_i) \right]^2$$

Classiquement, résoudre ce problème de minimisation aux moindres carrés revient à résoudre le système linéaire suivant :

$$Av = b \quad \text{avec} \quad A = \begin{bmatrix} \nabla E(p_1) \\ \nabla E(p_2) \\ \dots \\ \nabla E(p_{NxN}) \end{bmatrix} \quad \text{et} \quad b = \begin{bmatrix} -E_t(p_1) \\ -E_t(p_2) \\ \dots \\ -E_t(p_{NxN}) \end{bmatrix}$$

On résoud ainsi le problème de l'ouverture en extrayant le plus d'équations possibles pour un même pixel central à partir d'une fenêtre 5x5 centrée sur ce pixel. Avec les contraintes précédentes, on a 25 équations par pixel. Avec $[u \ v]^T$ le vecteur déplacement v_Q , on peut écrire :

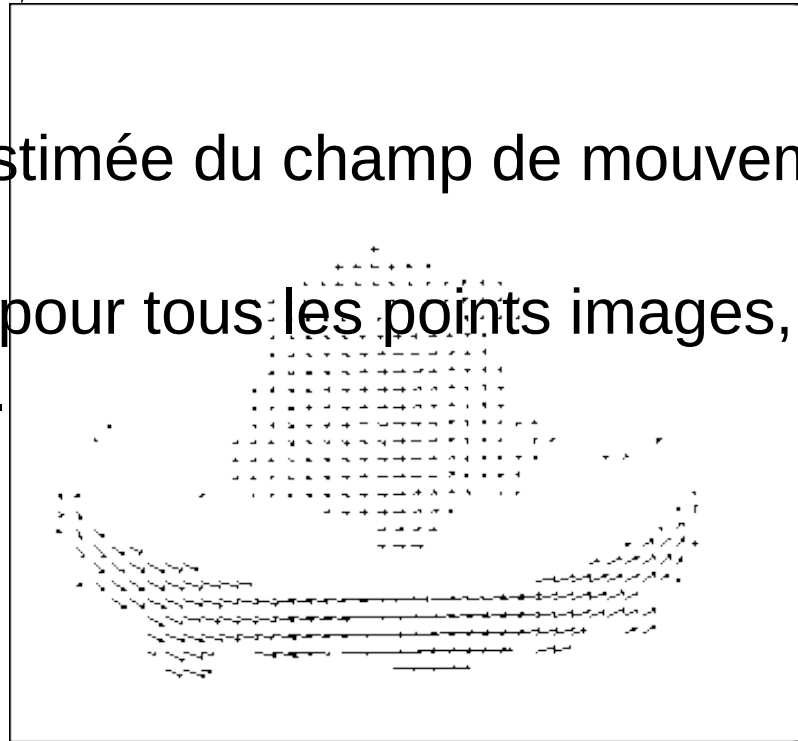
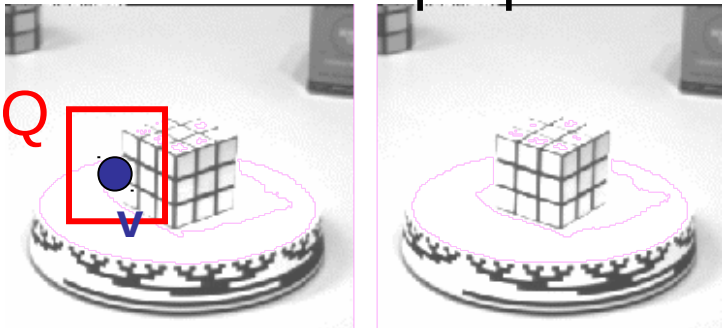
$$\begin{bmatrix} E_x(p_1) & E_y(p_1) \\ E_x(p_2) & E_y(p_2) \\ \vdots & \vdots \\ E_x(p_{25}) & E_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} E_t(p_1) \\ E_t(p_2) \\ \vdots \\ E_t(p_{25}) \end{bmatrix}$$

La solution de ce système est donnée par :

$$v = \left(A^T A \right)^{-1} A^T b$$

Où v est le flot optique (ou estimée du champ de mouvement) au centre du patch Q .

En répétant cette procédure pour tous les points images, on obtient un flot optique dense.

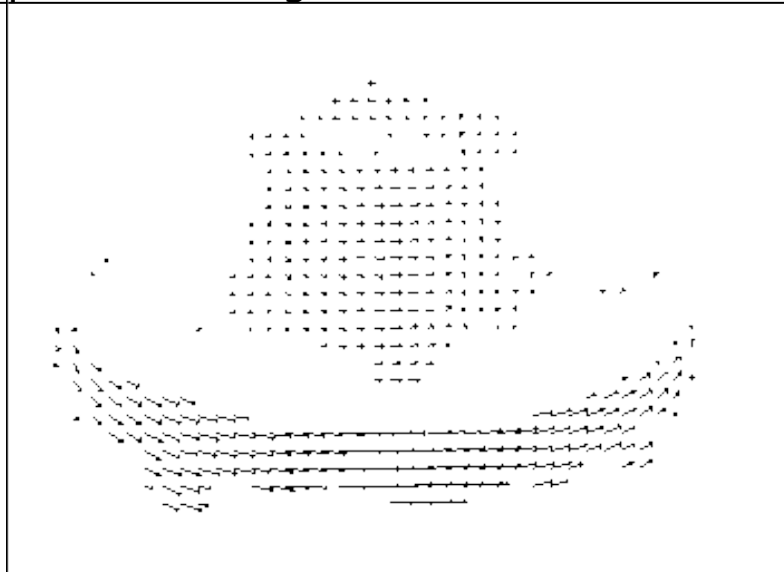


Algorithme CONSTANT_FLOW

INPUT : Une séquence temporelle de n images E_1, E_2, \dots, E_n . Soit Q une région carrée de $N \times N$ pixels (disons 5×5)

- Filtrer chaque image de la séquence avec un filtre Gaussien de déviation standard σ_s (disons $\sigma_s = 1,5$ pixels) le long de chaque dimension spatiale.
- Filtrer chaque image de la séquence le long de la dimension temporelle avec un filtre Gaussien de déviation standard σ_t (disons $\sigma_t = 1,5$ frames). Si $2k+1$ est la taille du filtre temporel, il faut laisser de côté les k premières et dernières images.
- Pour chaque pixel p de chaque image de la séquence :
 - Calculer la matrice A et le vecteur b
 - Calculer le flot optique $v(p) = (A^T A)^{-1} A^T b$

OUTPUT : le flot optique de la séquence d'images



Conclusions sur cette méthode de flot optique

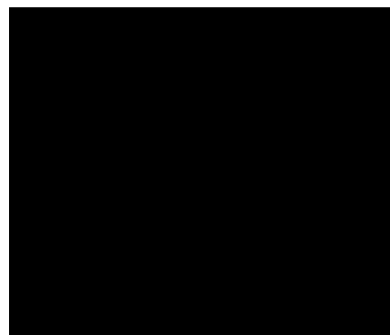
En quels pixels l'algorithme précédent éprouve-t-il des difficultés d'estimation ? A chaque fois que $A^T A$ est singulière :

$$A^T A = \begin{pmatrix} \sum E_x^2 & \sum E_x E_y \\ \sum E_x E_y & \sum E_y^2 \end{pmatrix}$$

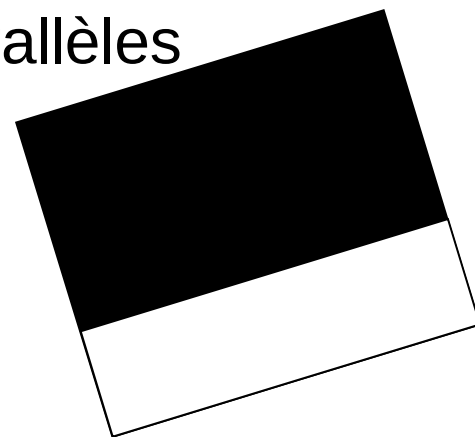
$A^T A$ caractérise la structure des niveaux de gris dans le patch Q . Ses vecteurs propres codent les directions principales des contours et ses valeurs propres la force de ces contours.

La matrice est singulière si au moins une des deux valeurs propres est nulle, ce qui correspond aux 2 cas suivants :

1. Tous les gradients spatiaux dans Q sont nuls



1. Tous les gradients spatiaux dans Q sont parallèles



Dans ce cas, le problème de l'ouverture ne peut pas être résolu et la seule possibilité est de retenir la solution de norme minimale $v(p) = (A^T A)^+ A^t b$ en s'aidant de la pseudo inverse et de la décomposition SVD.

Dans ce cas, on estime le flot normal.

Dernière remarque : l'algorithme précédent donne de bons résultats car la structure spatiale d'un champ de mouvement rigide est bien décrite par un polynôme de bas degré en les coordonnées images. Pour cette raison, l'hypothèse de conservation locale du champ de mouvement sur de petits patches images est raisonnable.

Techniques Différentielles : calcul du Flot Optique Itérative par Equa. Dif. : Horn et Shunck

Nous allons exposer une deuxième méthode utilisant la résolution de l'équation différentielle globale de conservation de l'illumination.

Soit la fonction 3D $f(x,y,t)$ représentant la valeur du niveau de gris en (x,y) à l'instant t . D'après ce qui précède on a,

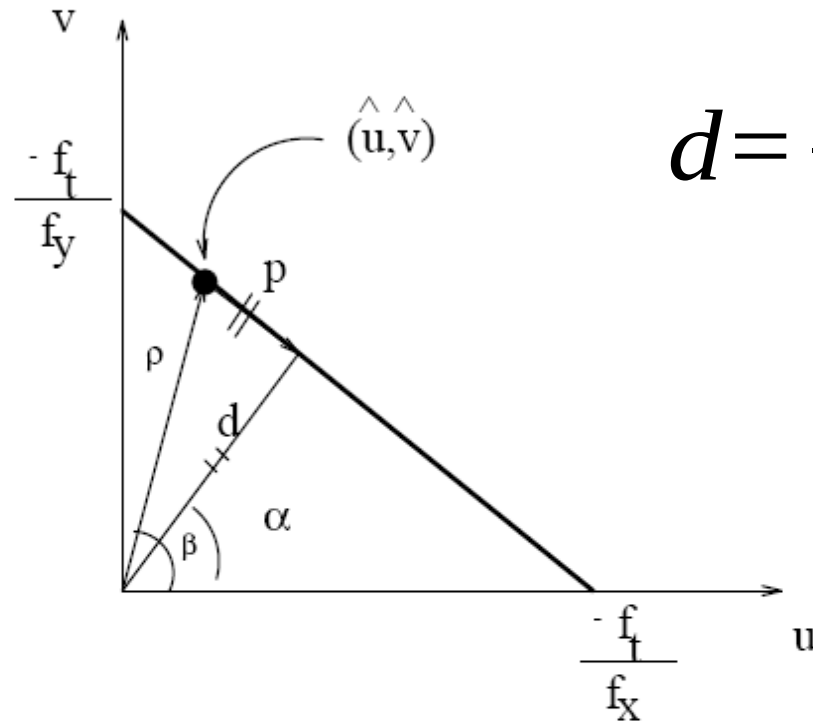
$$\left(\nabla f \right)^T \begin{bmatrix} u \\ v \end{bmatrix} + f_t = 0 \quad \text{Ce qui s'écrit :}$$

$$f_x u + f_y v + f_t = 0$$

C'est une équation linéaire à deux inconnues (u,v) constituant le flot optique car f_x , f_y et f_t sont calculables par des techniques de gradient classiques de l'imagerie 2D

La solution peut se situer n'importe où sur la droite :

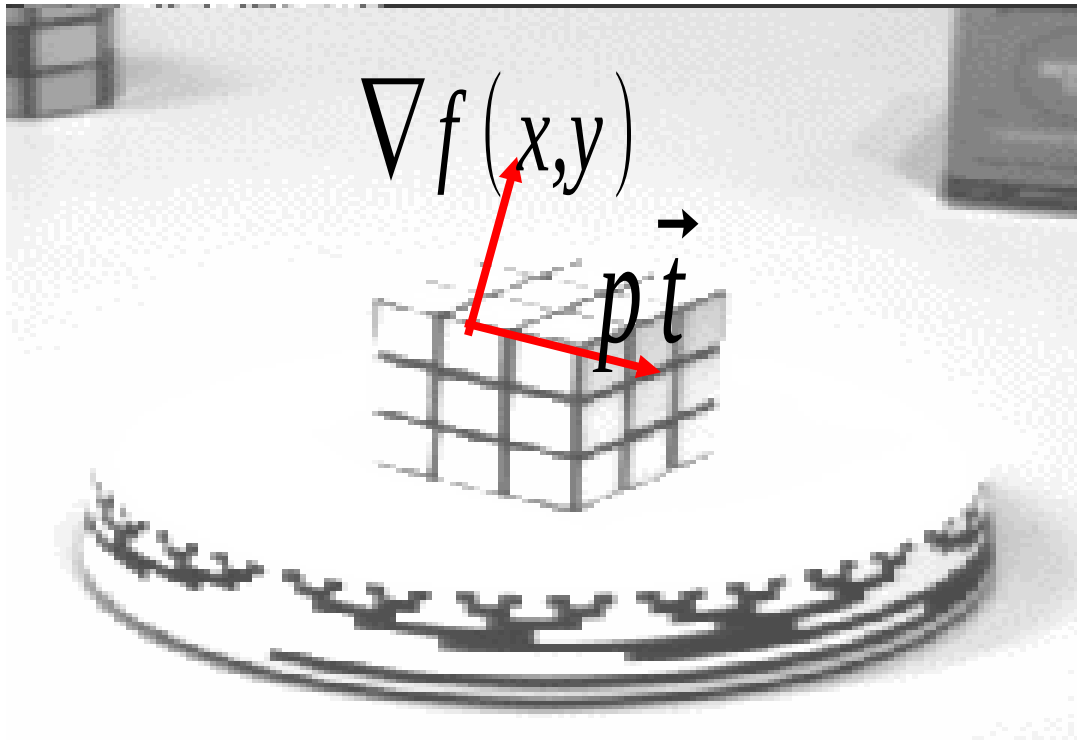
$$v = -\frac{f_x}{f_y}u - \frac{f_t}{f_y}$$



$$d = \frac{f_t}{\sqrt{f_x^2 + f_y^2}}$$

Optical flow constraint line in $u - v$ space. d is the length of perpendicular from the origin to the line, α is the angle the perpendicular makes with the x -axis. (\hat{u}, \hat{v}) is one possible solution. This optical flow vector, (\hat{u}, \hat{v}) , can be divided into two components: p , which is along the constraint line, and d which is perpendicular to the constraint line.

Cette valeur de d donne la composante du flot optique dans la direction du gradient d'intensité :



p ?

On ajoute une contrainte : le flot optique doit être lisse ou continu :

- On fait donc l'hypothèse que des points voisins sur une surface rigide possèdent des vecteurs déplacement localement approximativement identiques.

Cela revient à minimiser la fonctionnelle d'erreur suivante :

$E(u(x,y),v(x,y))$ pour simplifier :

$E(x,y)$ = terme d'attache aux données + contrainte de régularisation

$$E(x,y) = (f_x u + f_y v + f_t)^2 + \text{norme sur } [u,v]^T$$

$$E(x,y) = (f_x u + f_y v + f_t)^2 + \lambda (\|\nabla u(x,y)\|^2 + \|\nabla v(x,y)\|^2)$$

$$E(x,y) = (f_x u + f_y v + f_t)^2 + \lambda (u_x^2 + u_y^2 + v_x^2 + v_y^2)$$

Le paramètre λ code la force de la contrainte de lissage imposée.

https://tel.archives-ouvertes.fr/tel-00379735/file/these_ferreol.pdf

Technique classique dite de **régularisation de problème inverse**.

<https://www.ipol.im/pub/art/2012/g-cv/>

$$\text{or } \frac{\partial E}{\partial u} = 2(f_x u + f_y v + f_t) f_x + 2\lambda \left[\left(\frac{\partial}{\partial u} \frac{\partial u}{\partial x} \right) \frac{\partial u}{\partial x} + \left(\frac{\partial}{\partial u} \frac{\partial u}{\partial y} \right) \frac{\partial u}{\partial y} \right] = 0$$

$$\text{d'où } \frac{\partial E}{2\partial u} = (f_x u + f_y v + f_t) f_x + \lambda (u_{xx} + u_{yy}) = 0$$

$$\frac{\partial E}{2\partial v} = (f_x u + f_y v + f_t) f_y + \lambda (v_{xx} + v_{yy}) = 0$$

soit

$$(f_x u + f_y v + f_t) f_x + \lambda (\Delta u) = 0$$

$$(f_x u + f_y v + f_t) f_y + \lambda (\Delta v) = 0$$

Classiquement, en utilisant les masques de convolution linéaire, on peut calculer le laplacien d'une fonction par la formule suivante :

$$\Delta u = u - u_{\text{moyen}} \quad h = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

où u_{moyen} est la moyenne de la composante u du flot optique pris parmi les 4 voisins du pixel. Soit

$$(f_x u + f_y v + f_t) f_x + \lambda (u - u_{\text{moyen}}) = 0$$

$$(f_x u + f_y v + f_t) f_y + \lambda (v - v_{\text{moyen}}) = 0$$

On peut ainsi calculer (u,v) de façon itérative avec les équation :

$$u = u_{\text{moyen}} - f_x \frac{f_x u_{\text{moyen}} + f_y v_{\text{moyen}} + f_t}{\lambda + f_x^2 + f_y^2} = u_{\text{moyen}} - f_x \frac{P}{D}$$

$$v = v_{\text{moyen}} - f_y \frac{P}{D}$$

Algorithme HORN_FLOW

INPUT : Une séquence temporelle de n images f_1, f_2, \dots, f_n .

- Filtrer chaque image de la séquence avec un filtre de Dérivation le long de chaque dimension spatiale.
- Filtrer chaque image de la séquence le long de la dimension temporelle avec un filtre de Dérivation.

1. Initialisation : Pour chaque pixel p de chaque image de la séquence :

- $k=0$
- Initialise u^k et v^k à zéro

2. Pour chaque pixel p de chaque image de la séquence :

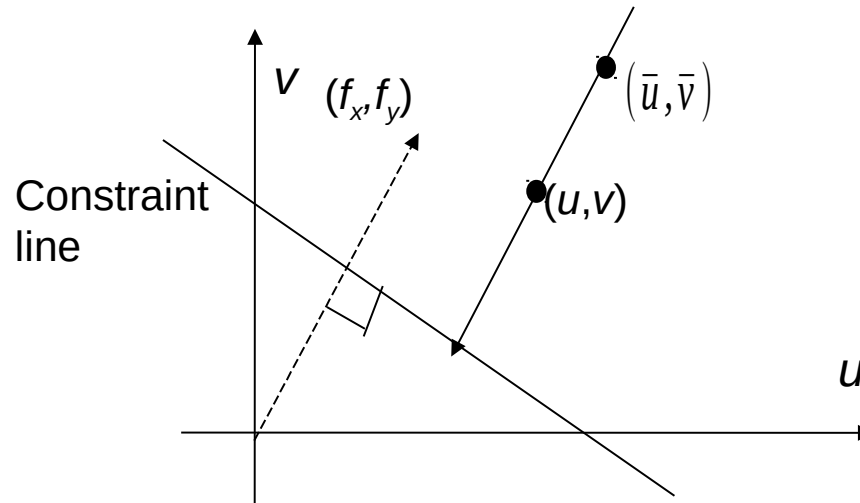
- Jusqu'à ce qu'une mesure d'erreur soit satisfaite, faire :

$$u^k = u_{moyen}^{k-1} - f_x \frac{P^{k-1}}{D}$$

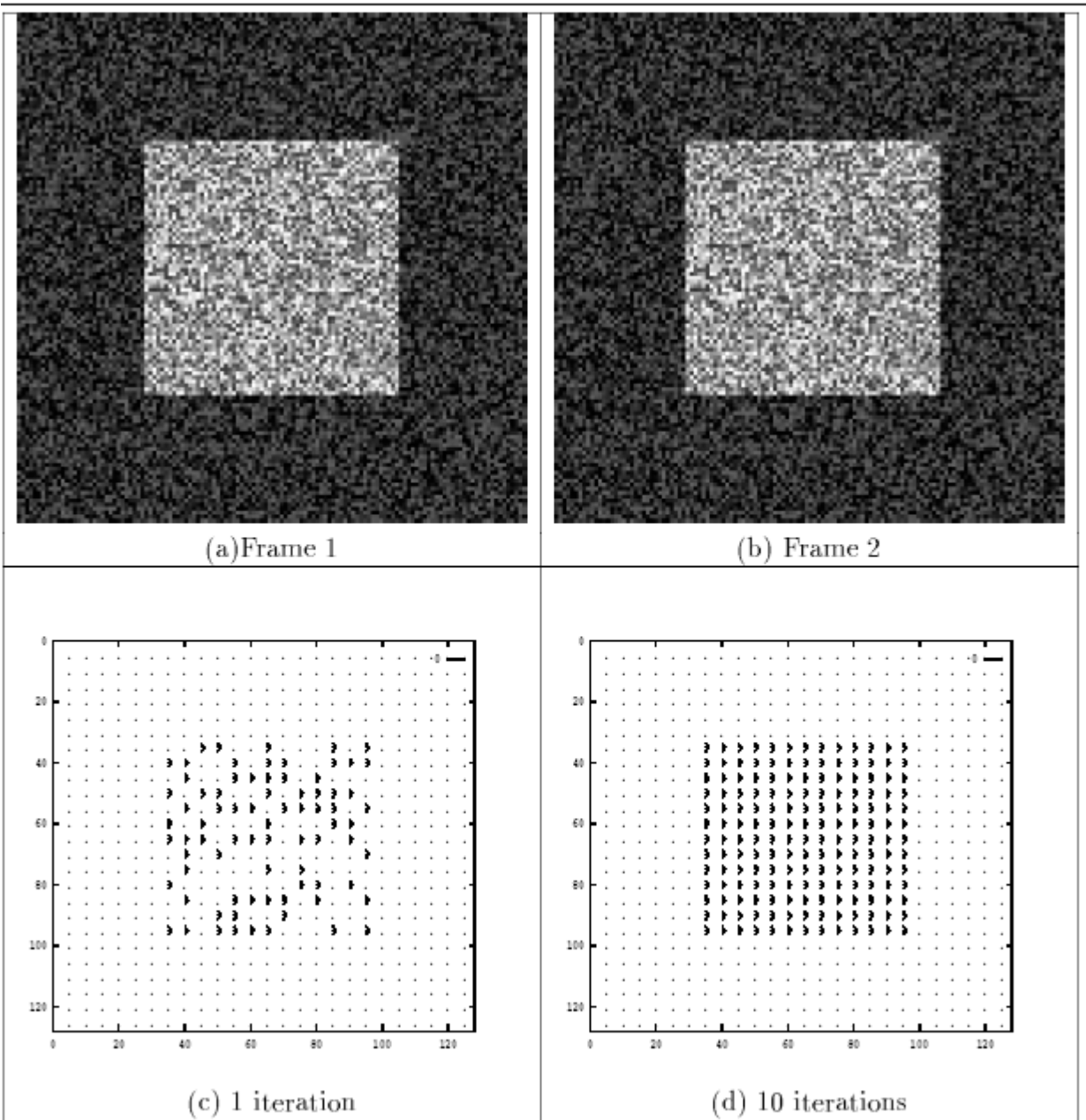
$$v^k = v_{moyen}^{k-1} - f_y \frac{P^{k-1}}{D}$$

OUTPUT : le flot optique de la séquence d'images

Interprétation Intuitive du Schéma Itératif



La nouvelle valeur de (u, v) en un point (x, y) est égale à la moyenne des valeurs voisines moins un ajustement dans la direction du gradient d'intensité



Results for Horn and Schunck algorithm for displacement of 1 pixel and $\lambda = 4$.

Schunk propose d'ajouter d'autres contraintes géométriques à celle utilisée précédemment .

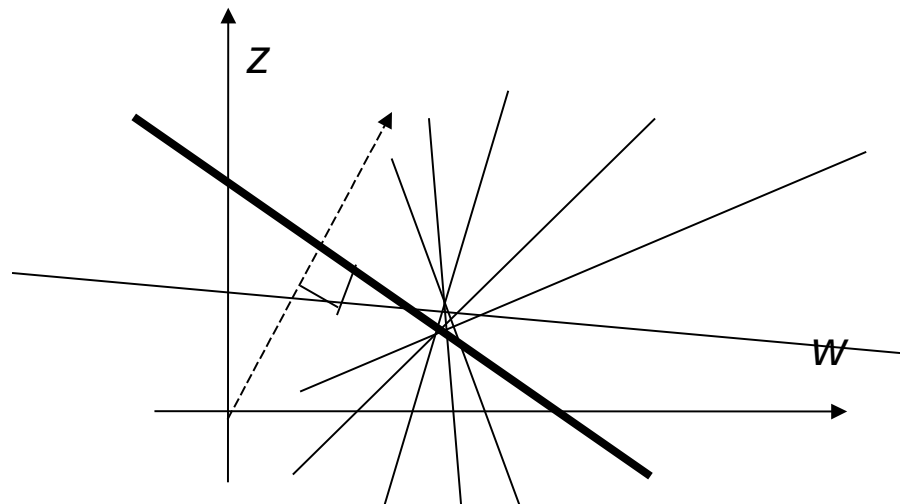
Le niveau de gris en un pixel donne une contrainte, à savoir que le flot optique peut se trouver n'importe où sur la droite déterminée par les dérivées spatiales et temporelles.

Si on utilise un pixel avoisinant, la même contrainte appliquée à ce pixel fournit une nouvelle droite de contrainte.

Le flot optique est ainsi l'intersection de ces deux droites de contrainte.

On ne peut pas privilégier un voisin, on considère donc les 8 contraintes générées par les 8 pixel du voisinage 3x3 en 8 connexité.

Faisceau de
droites de
contraintes



Bien sûr, ces droites ne sont pas exactement concourantes, même si leurs points d'intersection se regroupent dans une région étroite.

Il s'agit d'estimer le meilleur point d'intersection.

Passons en coordonnées polaires (ρ, β) :

$$d = \rho \cos(\alpha - \beta)$$

$$\text{avec } d = \frac{f_t}{\sqrt{f_x^2 + f_y^2}} \quad \text{et} \quad \alpha = \arctan \frac{f_y}{f_x}$$

ρ est la vitesse et β la direction du flot optique

2 points 1 et 2 fournissent les 2 équations de flot optique suivante :

$$d_1 = \rho \cos(\alpha_1 - \beta)$$

$$d_2 = \rho \cos(\alpha_2 - \beta)$$

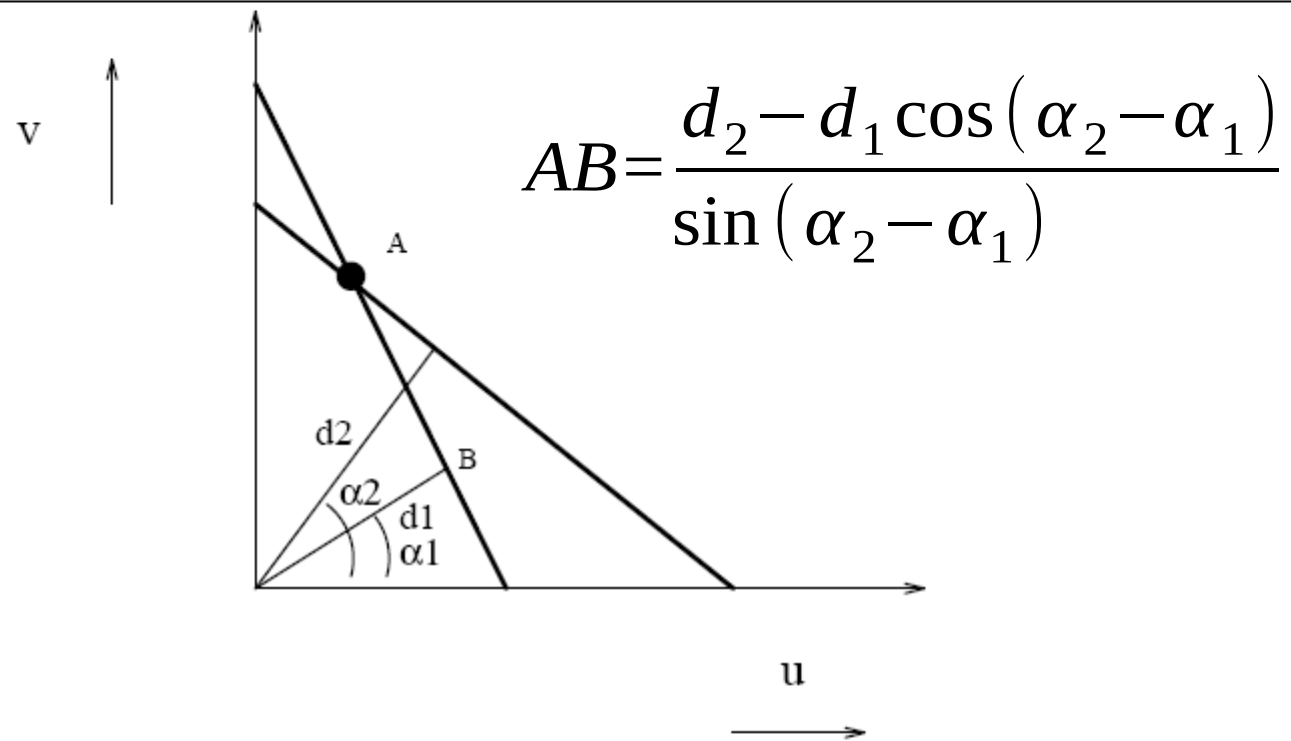


Figure 5.5: Intersection of two optical flow constraint lines.

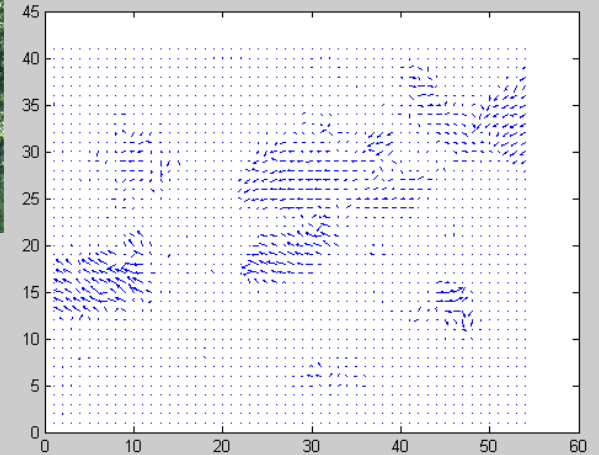
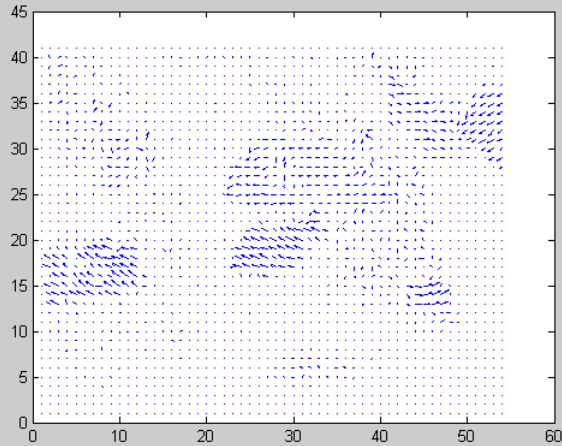
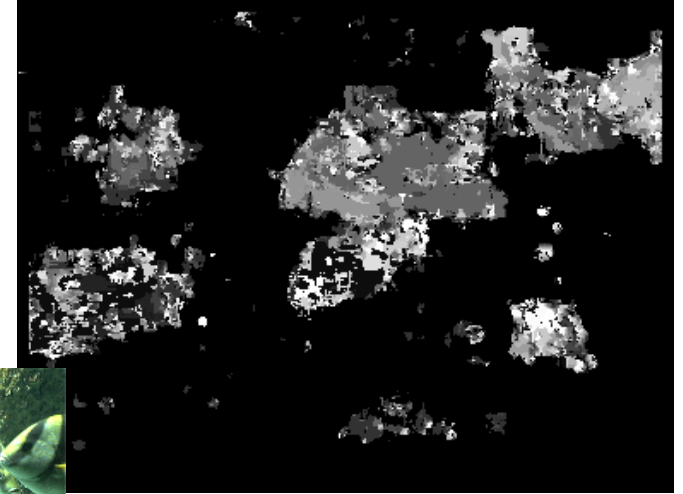
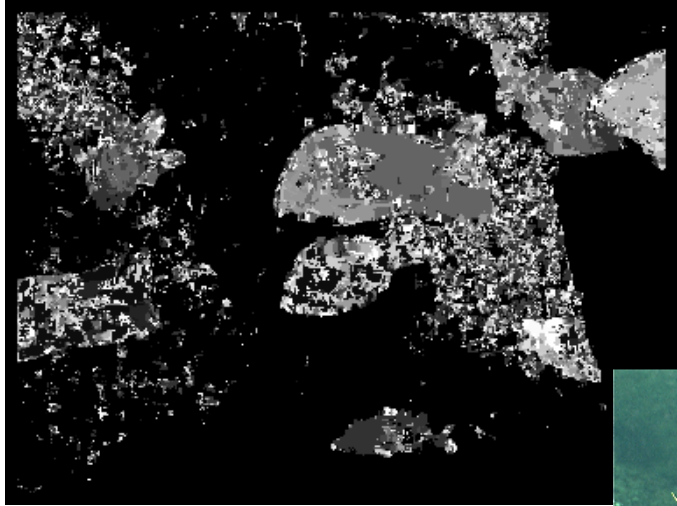
De façon similaire, les intersections avec les droites correspondant aux points 3, 4, ..., 9 peuvent être calculées.

Ces intersections sont sensés se regrouper autour d'un point central.

Si l'on désigne par \hat{b} ce point central, alors le flot optique est donné par :

$$\rho = \sqrt{d_1^2 + \hat{b}^2}$$

$$\beta = \alpha_1 + \arctan\left(\frac{\hat{b}}{d_1}\right)$$



Représentation des normes des vecteurs vitesses et du flot optique évalués par appariement à partir d'images en niveaux de gris à gauche, à partir d'images en couleurs à droite

$$\left\{ v_x(x,y) = a_{x0} + a_{xx}x + a_{xy}y \right\}$$

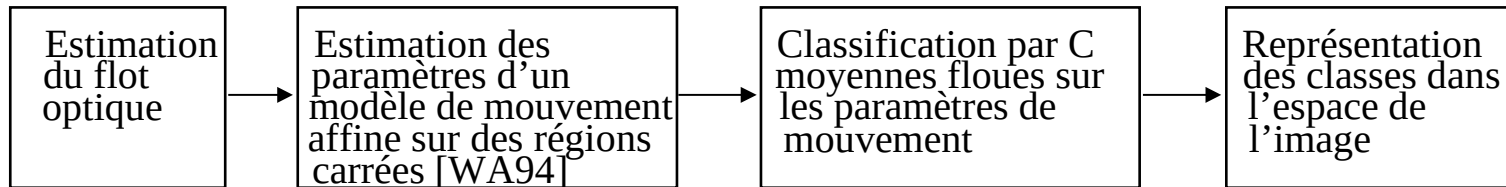


image initiale

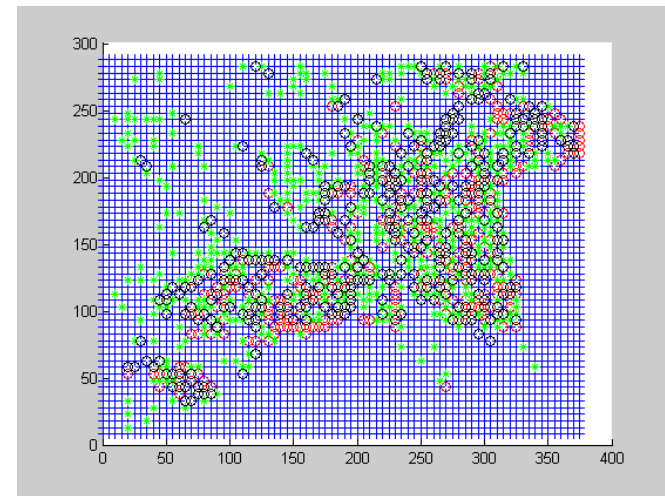
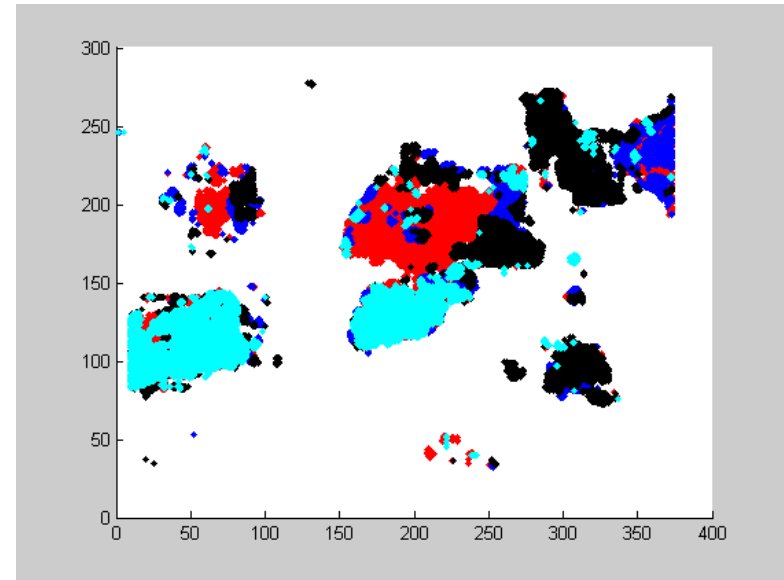


image segmentée



image initiale



segmentation par classification sur
les composantes de la vitesse

<https://helios2.mi.parisdescartes.fr/~lomn/Cours/CV/SeqVideo/CoursComplements/Recherche/Stage2002SegmMouvement.ppt>

Voir wang1.pdf et wang2.pdf dans CoursComplements.

<https://www.youtube.com/watch?v=v6cGD2iE548>

Bibliographie

- “Performance of Optical Flow Techniques”, J. Barron et al., IJCV, vol. 12, pp. 43-77, 1994
- “Determining Optical Flow”, Horn et Schunck, AI, vol.17, pp. 185-203, 1981