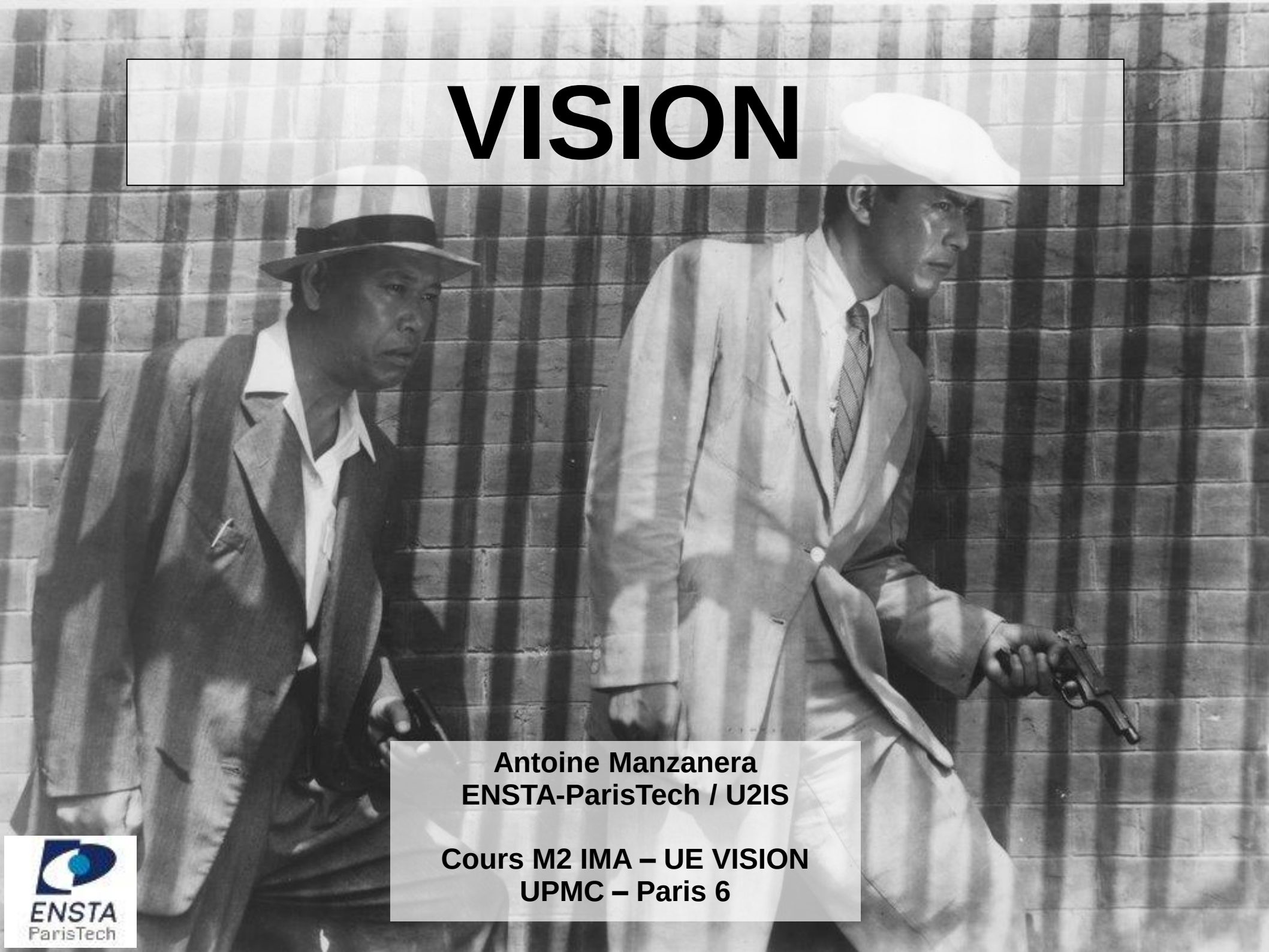
VISION



Antoine Manzanera
ENSTA-ParisTech / U2IS

Cours M2 IMA – UE VISION
UPMC – Paris 6

ENSTA ParisTech

VISION COURS N°2 :
Quelques approches de co-conception pour l'analyse du mouvement et la reconstruction 3d par vision artificielle

Objectifs du cours :

- Compléter les approches classiques (Algorithmique / Stéréovision / Structure from Motion) de reconstruction 3d et d'analyse du mouvement par les approches de co-conception, qui tirent parti de l'ensemble des éléments optiques / mécaniques / électroniques d'un système pour accroître ses capacités de perception et d'analyse.

- Comprendre le principe des différentes approches de co-conception de systèmes de vision par ordinateur.

VISION COURS N°2 :
Quelques approches de co-conception pour l'analyse du mouvement et la reconstruction 3d par vision artificielle

- **1ère Partie, 3d actif :**
  - caméras temps de vol
  - lumière structurée.
- **2ème Partie, 3d passif :**
  - caméra plénoptique
  - profondeur par le focus
  - ouverture codée
- **3ème Partie, Rétines électroniques :**
  - Rétines dédiées à l'analyse du mouvement
  - Rétines programmables

1ère Partie : CAMERAS 3D / APPROCHES ACTIVES

Les caméras 3d « actives » cherchent à mesurer la profondeur de tous les points projetés sur le plan image à partir de la réponse qu'il fournissent à un éclairage particulier.

Leurs deux composantes fondamentales sont donc :

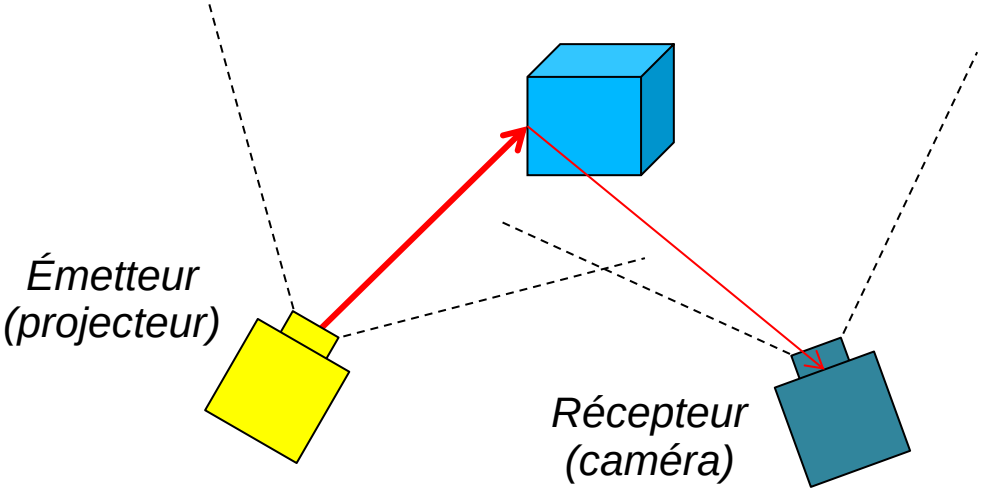
1. Un système d'illumination contrôlé dans le temps et dans l'espace
2. Un système chargé d'analyser l'image de la scène illuminée

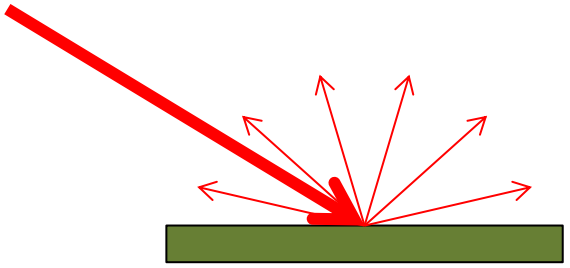
Ces systèmes sont actifs dans le sens qu'ils *émettent* un signal lumineux (à ne pas confondre avec le sens habituelle de la « vision active », i.e. qui se « déplace pour voir »).

APPROCHES ACTIVES ET MODELES DE DIFFUSION

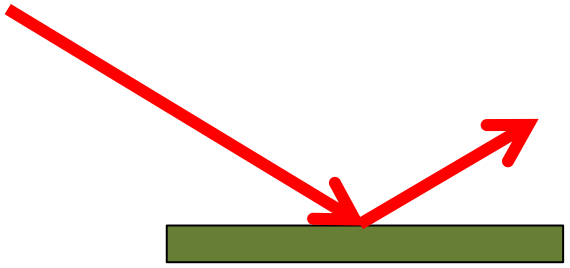
Pour les caméras 3d actives, on suppose que tout point illuminé par le projecteur dans le champ de la caméra réfléchit une partie de sa lumière reçue vers le centre optique de la caméra.

Le type de diffusion du point mesuré joue donc un rôle majeur dans l'estimation de sa profondeur…

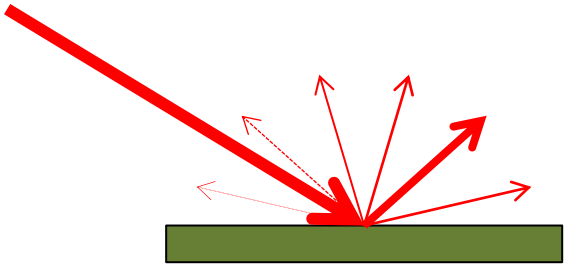




Diffusion lambertienne



Réflexion spéculaire parfaite (miroir)



Diffusion partiellement spéculaire

Remarque : ce problème est également important pour les approches passives (ex : appariement de structures d'une pose à l'autre).

3D ACTIF : CAMERA “TEMPS DE VOL”

Les caméras 3d « temps de vol » (ToF) mesure la distance d_x d’un point X projeté en x, à partir du temps t_x de propagation de la lumière (de vitesse c) depuis son émission par le projecteur jusqu’à sa réception par le photocapteur associé à x, après avoir été réfléchie par le point X :

$$d_x = \frac{c.t_x}{2}$$

Contrairement aux systèmes de type scanner (LIDAR), la lumière émise par les caméras ToF (en général laser ou LED infrarouge) illumine toute la scène simultanément.

Différentes technologies peuvent être utilisées pour mesurer le temps de vol :

- Mesure directe du temps (illumination impulsionnelle)
- Mesure de déphasage (illumination continue modulée dans le temps).

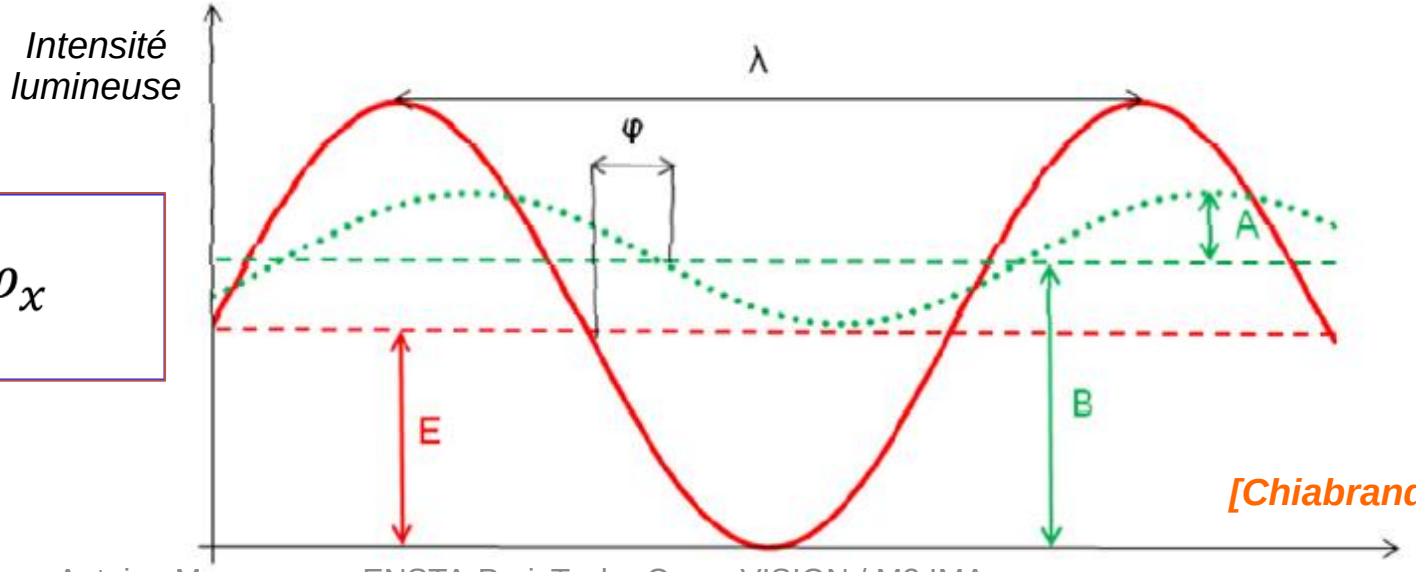
[CamCube - ©PMDTech]

[Kinect v2 for Xbox One - ©Microsoft]

3D ACTIF : CAMERA ToF PAR MESURE DE PHASE

- La scène est uniformément illuminée avec une intensité qui varie temporellement selon un signal sinusoïdal (en rouge) d'amplitude E.
- Le signal reçu en un pixel *x* (en vert) a la même fréquence, une amplitude A plus faible qui dépend de la réflectivité du point et un déphasage φ qui dépend de sa distance.
- Le signal reçu est décalé en intensité (offset) d'une valeur B à cause de la lumière de fond présente dans la scène.
- Le signal reçu est échantillonné et la phase φ est déduite des intensités mesurées.
- La période de modulation λ (typ. 50 ns) est grande par rapport au temps de vol pour éviter les ambiguïtés de phase, mais petite par rapport au temps de pose classiques pour permettre de répéter la mesure (filtrage temporel).

$$d_x = \frac{c\lambda}{4\pi}\varphi_x$$



[Chiabrando 2009]

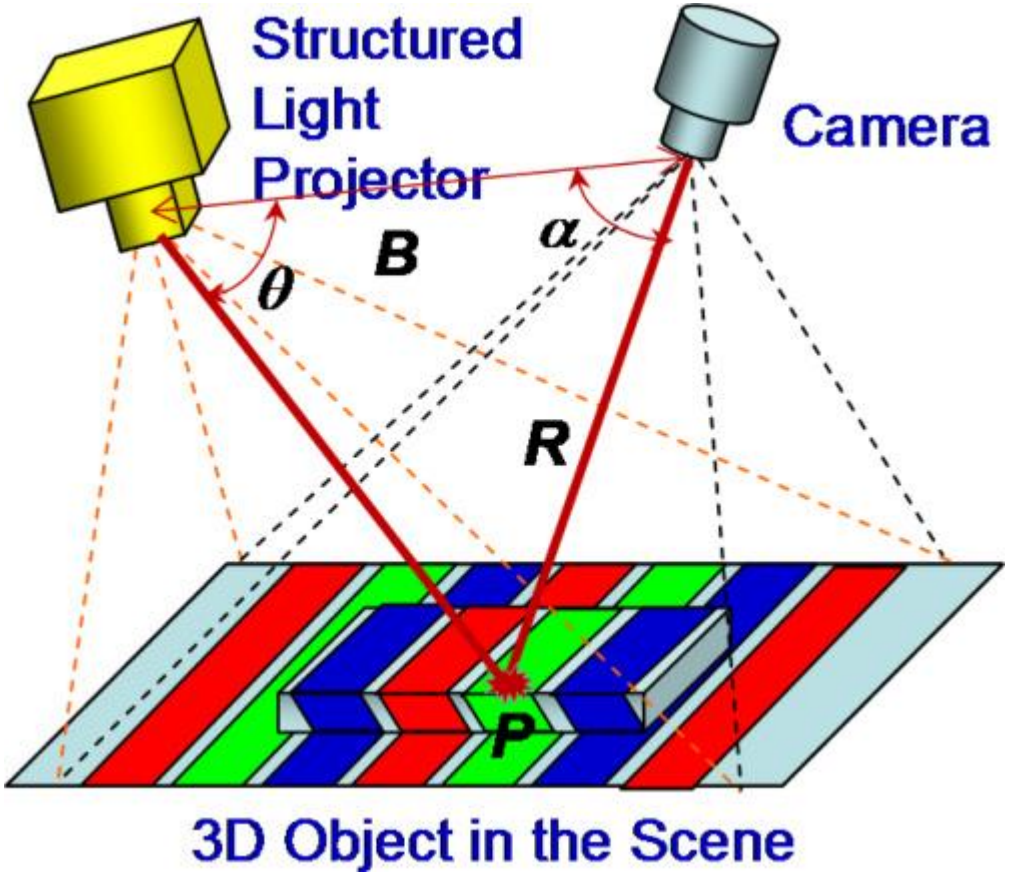
3D ACTIF : CAMERA “LUMIERE STRUCTUREE”

Les caméras 3d à lumière structurée interprètent la déformation d’une image 2d projetée dans la scène pour retrouver l’information de profondeur.

Elles se fondent sur le même principe de triangulation que la stéréovision :

$$R = B \frac{\sin\theta}{\sin(\alpha + \theta)}$$

La structure des images 2d projetées détermine une codification spatiale qui joue un rôle majeur dans la triangulation.

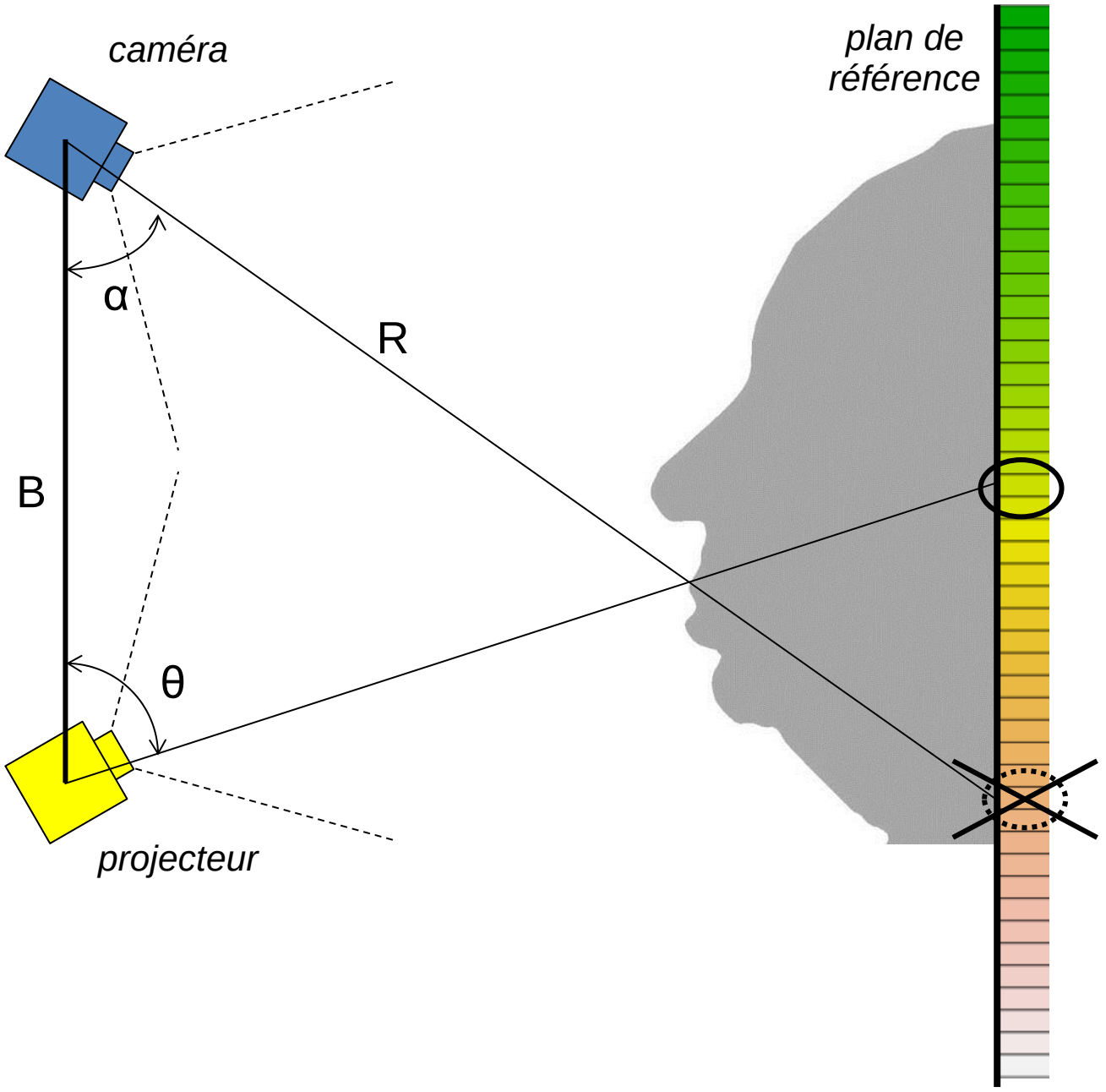


[Geng 2011]

3D ACTIF : CAMERA “LUMIERE STRUCTUREE”

$$R = B \frac{\sin\theta}{\sin(\alpha + \theta)}$$

L’angle α est fourni par la position du point dans l’image, et l’angle θ par la couleur (ou le motif) correspondant dans le plan de référence :



3D ACTIF : CAMERA “LUMIERE STRUCTUREE”

De même :

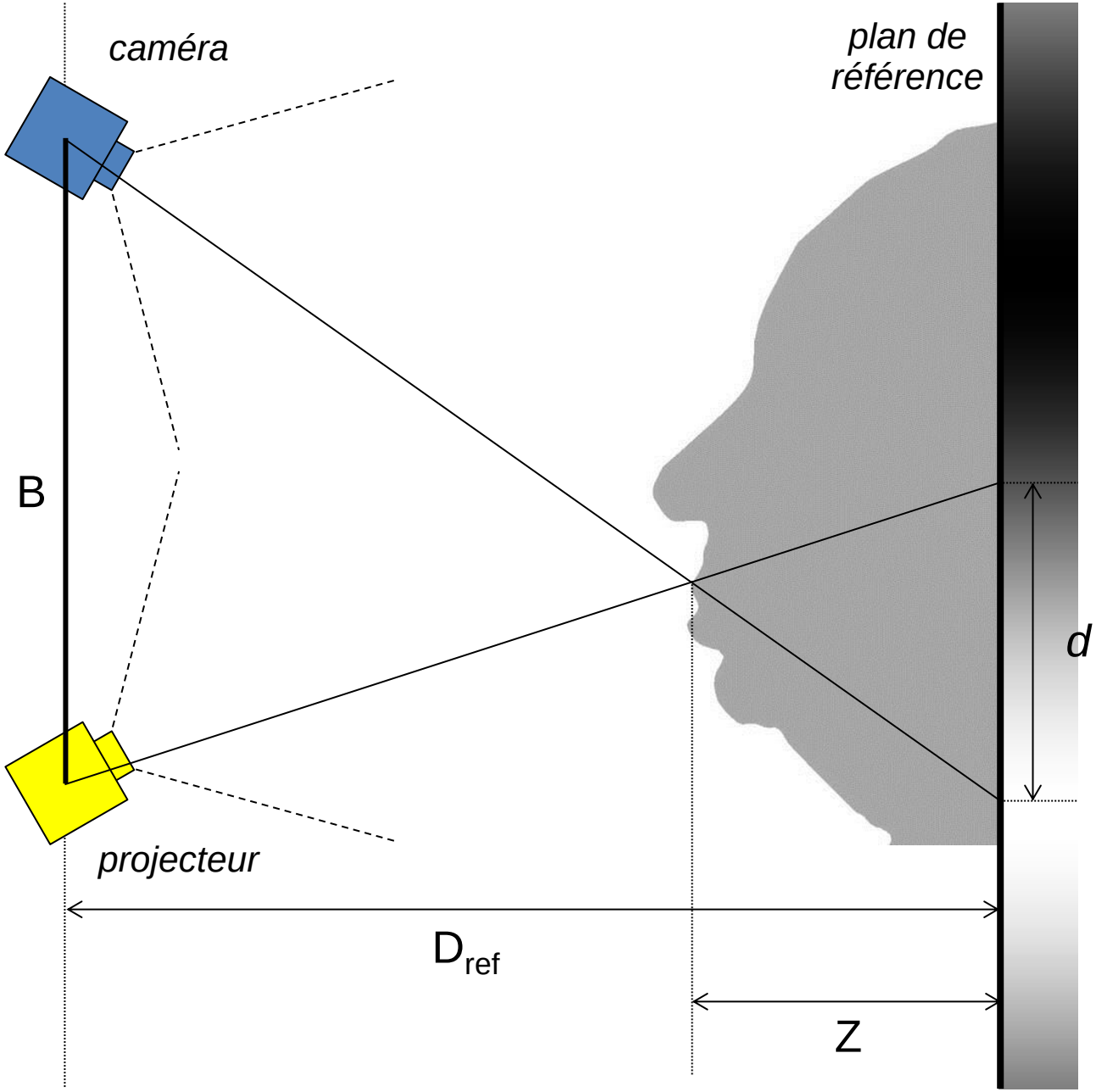
$$\frac{d}{B} = \frac{Z}{D_{ref} - Z}$$

Et donc :

$$Z \approx \frac{D_{ref}}{B} d$$

Donc, si l’image projetée est une rampe sinusoïdale, la profondeur peut se déduire de la différence de phase :

$$Z \propto \Delta\varphi$$



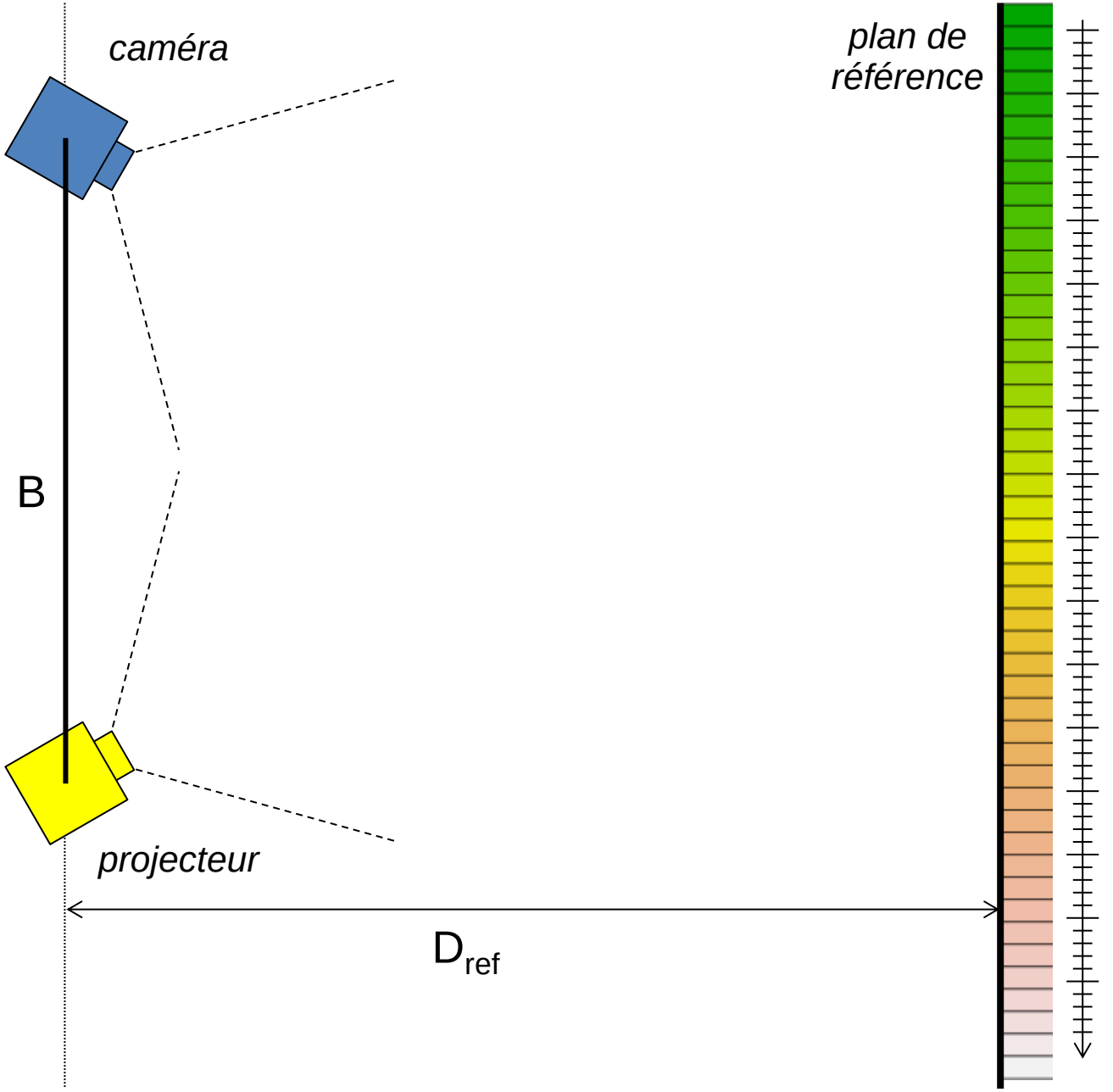
CAMERA “LUMIERE STRUCTUREE” : CALIBRATIONS

A l’instar des systèmes de stéréovision, la caméra et le projecteur doivent être calibrés de façon à :

(1) Déterminer la droite de rétroprojection de chaque pixel de l’image captée.

(2) Associer à chaque motif de l’image projetée une direction correspondant à la projection du point correspondant d’un plan de référence.

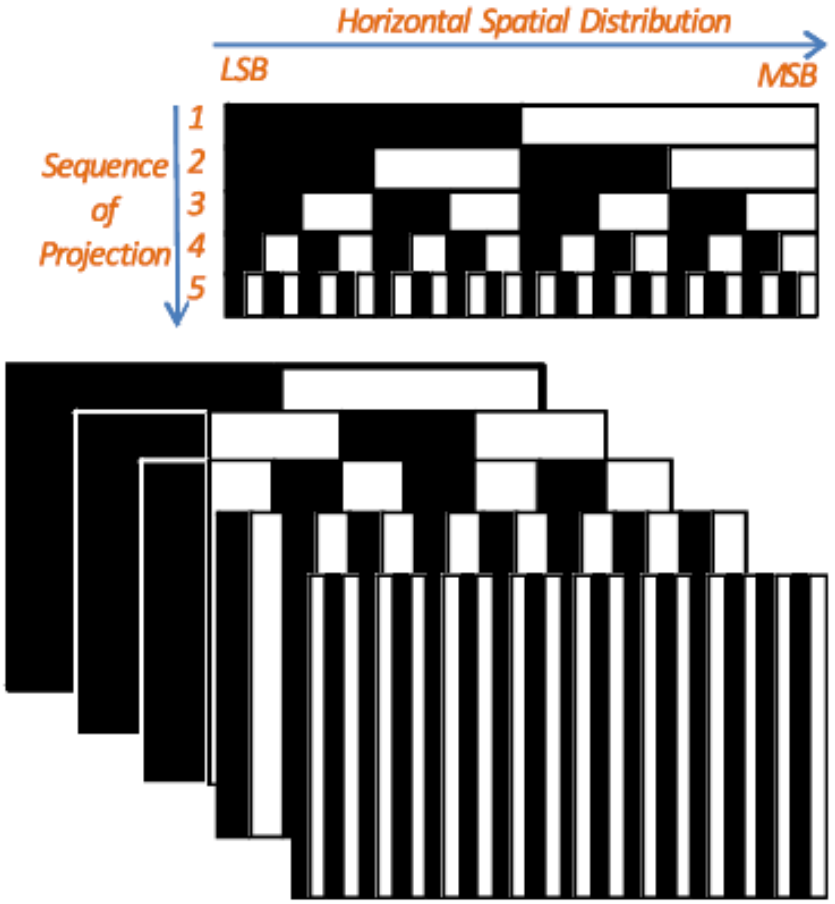
(Voir cours Calibration)



LUMIERE STRUCTUREE : QUELS MOTIFS PROJETER ?

- Idéalement, on voudrait identifier chaque point observé à partir de sa valeur / couleur…
  - …mais toutes les valeurs doivent se distinguer facilement !
- On peut aussi identifier un point à partir de son voisinage…
  - …mais alors chaque voisinage doit être unique !
- La profondeur étant associée à un angle, une mire 1d (bande) peut suffire…
  - ..mais utiliser une mire 2d peut permettre de résoudre des ambiguïtés !
- On peut aussi combiner plusieurs mires séquentiellement...
  - …mais alors le temps d'acquisition augmente !

LUMIERE STRUCTUREE : MIRES SEQUENTIELLES



Séquence binaire 2^5

[Posdamer 1982, tiré de Geng 2011]

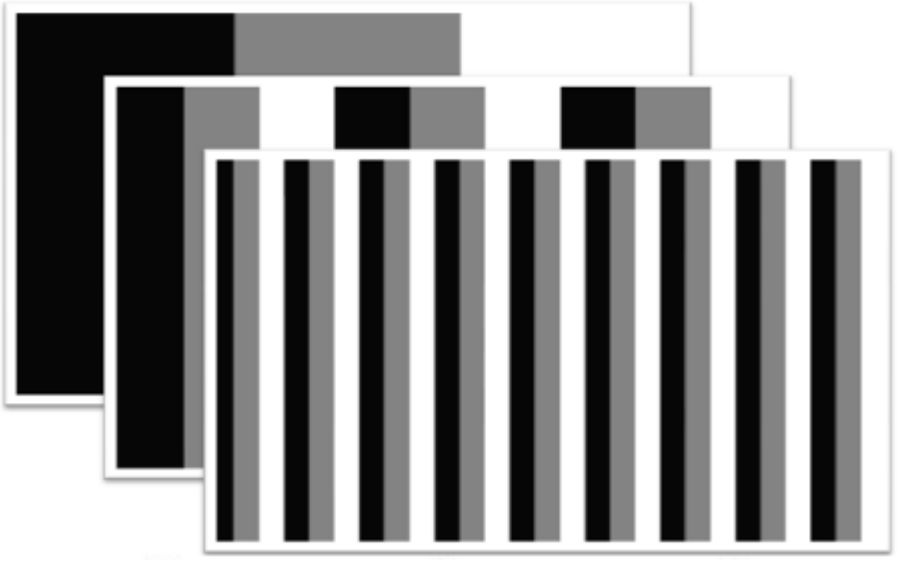
L'utilisation de mires binaires permet de discriminer les valeurs de façon optimale.

La résolution en profondeur dépend du nombre de valeurs finales, et donc pour les techniques séquentielles, de la durée de l'acquisition.

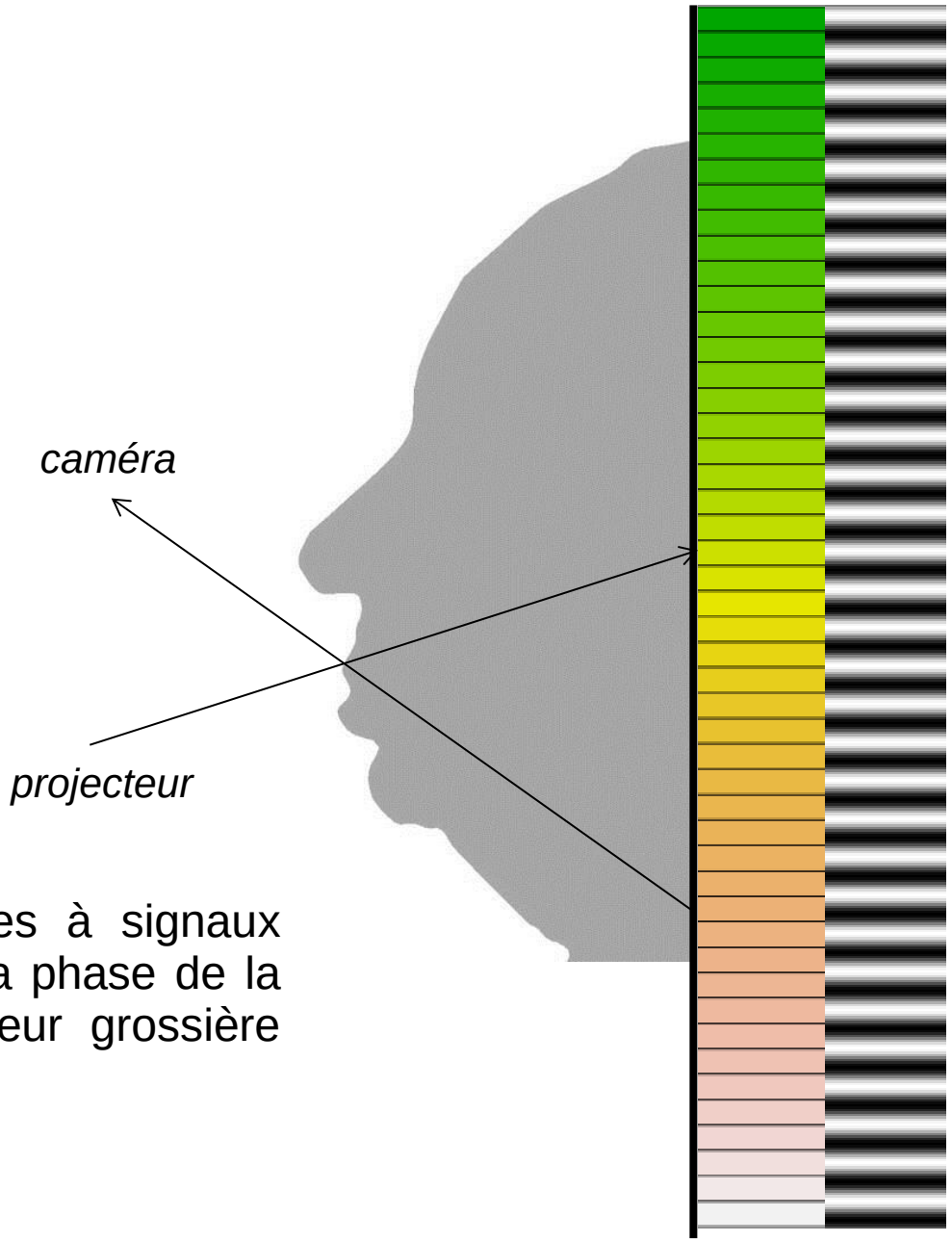
[tiré de Naramsimhan 2006]

LUMIERE STRUCTUREE : MIRES SEQUENTIELLES

Augmenter le nombre de bits (ci-dessous) : compromis contraste/temps d'acquisition.



Séquence ternaire 3^3

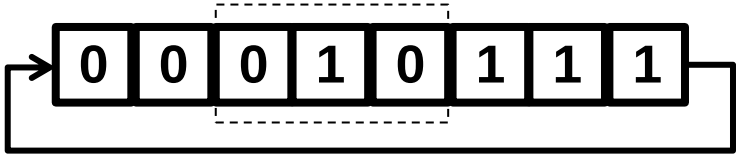


On peut aussi combiner (ci-contre) les mires à signaux rectangulaires avec les mires sinusoïdales : la phase de la mire sinusoïdale permet d'affiner la profondeur grossière estimée par la mire rectangulaire.

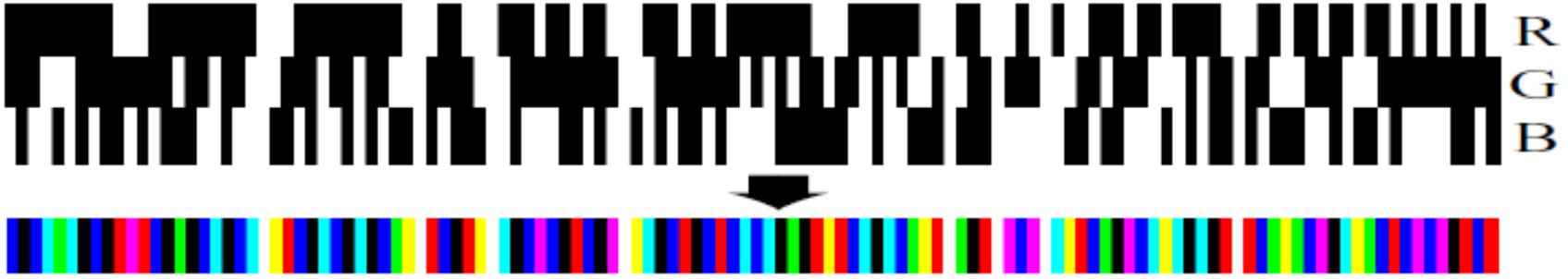
LUMIERE STRUCTUREE : MIRE UNIQUE “SNAPSHOT”

- Pour mieux distinguer les valeurs de la mire, les mires rectangulaires (plages) sont préférables aux mires continues (rampes).
- Pour pouvoir discriminer localement les points avec des valeurs quantifiées, on peut utiliser des motifs locaux (voisinages) au lieu de la valeur seule.
- Mais il faut alors que chaque motif définisse de façon *unique* une position.

Les séquences de De Bruijn *B(n,k)* sont des mots d’un alphabet à *n* symboles tels que tous les sous-mots de longueur k qu’on peut extraire sont différents.



Séquence de De Bruijn B(2,3)

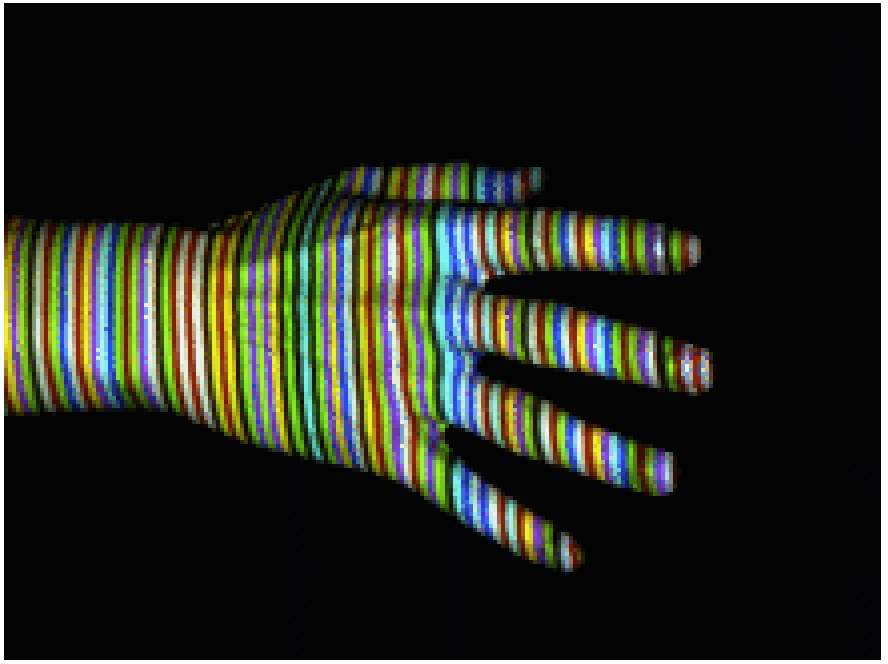


Séquence couleur de De Bruijn B(5,3)

[Zhang 2002]

LUMIERE STRUCTUREE : MIRE UNIQUE “SNAPSHOT”

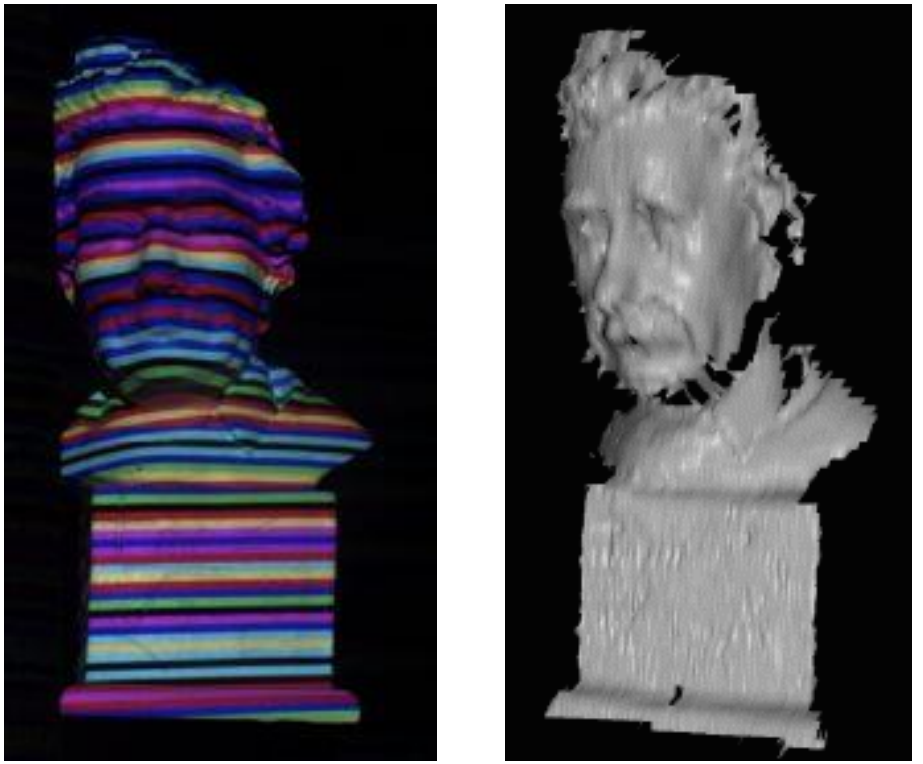
L’utilisation d’une mire unique (« snapshot ») permet de réduire considérablement le temps d’acquisition et donc d’acquérir des scènes mobiles :



[Zhang 2002]

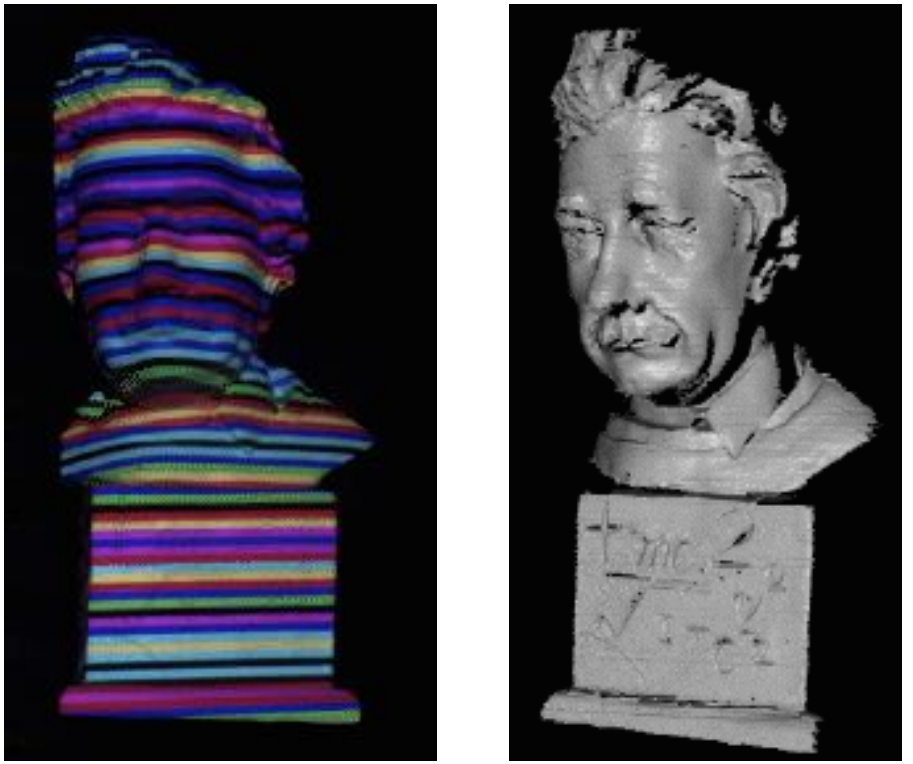
MIRE DE DE BRUIJN : SNAPSHOT *VS* SEQUENTIEL

Les mires conçues pour l'acquisition unique peuvent être utilisées *déphasées* pour des acquisitions séquentielles, de façon à améliorer la robustesse et la résolution (scènes statiques) :



Acquisition unique « snapshot »

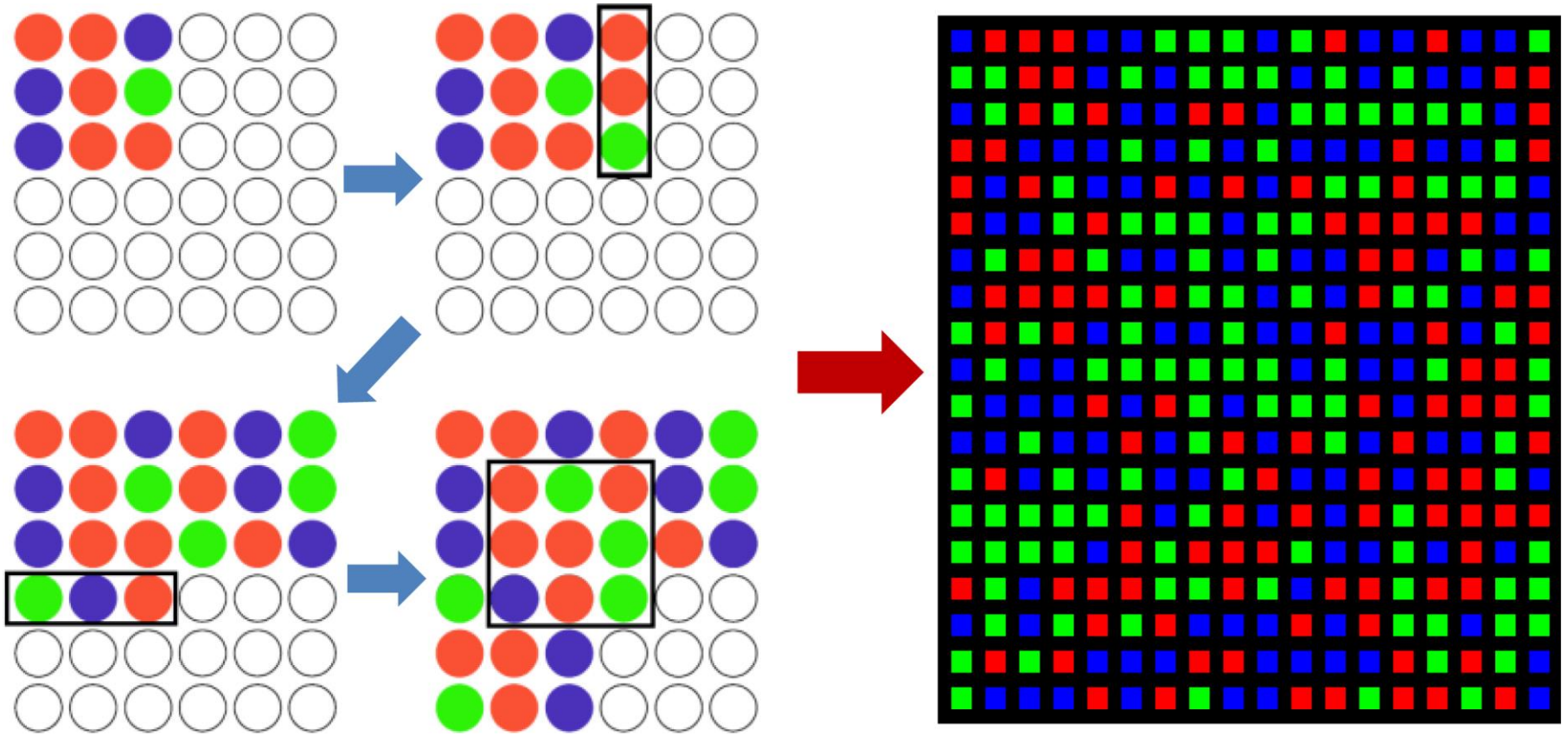
[Zhang 2002]



Acquisition séquentielle :
7 mires entrelacées

LUMIERE STRUCTUREE : MIRE UNIQUE “SNAPSHOT”

Mire 2d « snapshot » par motifs pseudo-aléatoires générés par algorithme « brute-force »:

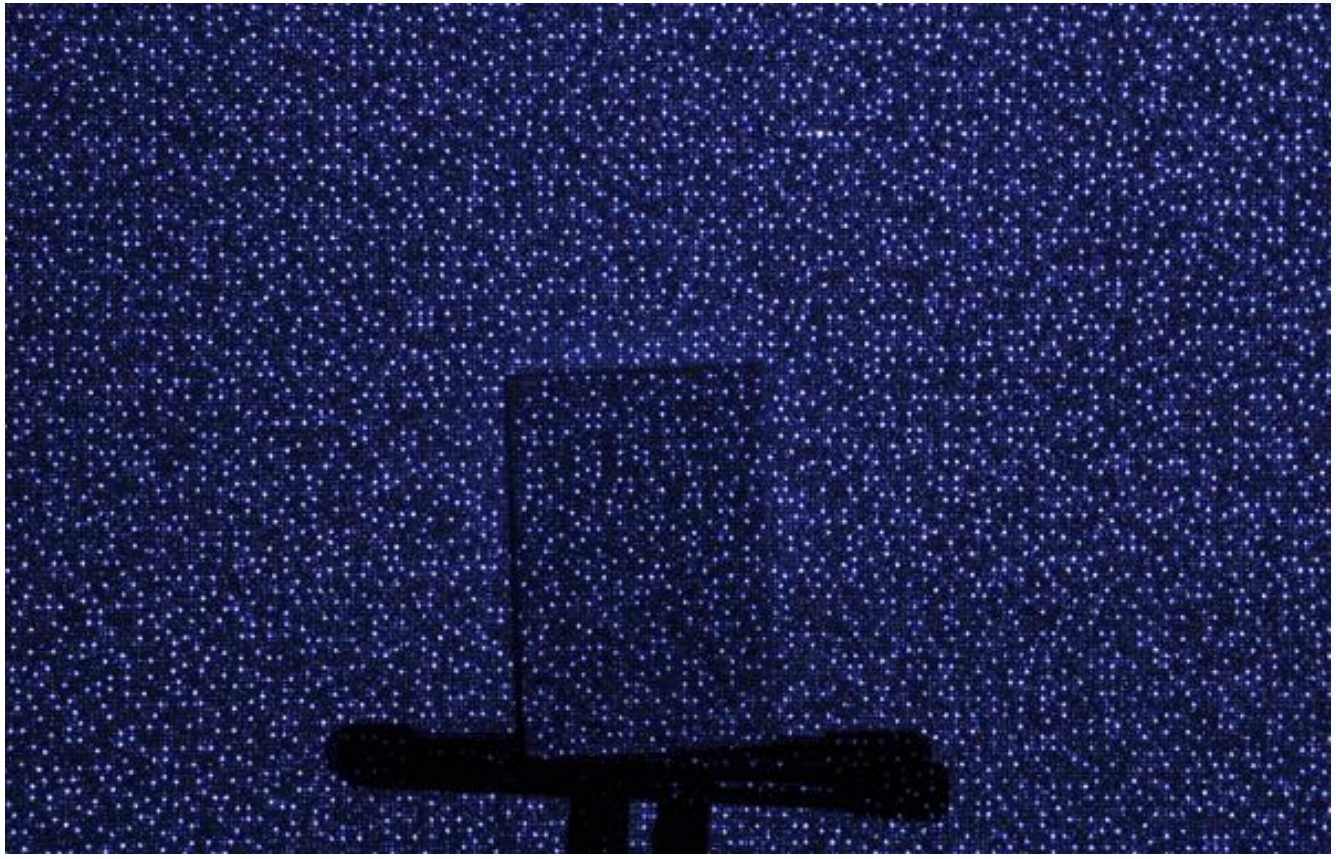


[Geng 2011]

LUMIERE STRUCTUREE : MIRE UNIQUE “SNAPSHOT”

La première version de la Kinect™ contient une caméra RGB associée à une caméra à lumière structurée qui utilise une mire pseudo-aléatoire infra-rouge.

[Kinect v1 - © Microsoft]



[© futurepicture.org]

2ème Partie : CAMERAS 3D / APPROCHES PASSIVES

Pour des raisons de consommation d'énergie et de discrétion, il est souvent préférable pour un système d'observation de ne pas émettre de signal lumineux.

Les techniques passives récupèrent l'information utile uniquement à partir de l'intensité lumineuse reçue par les photodétecteurs.

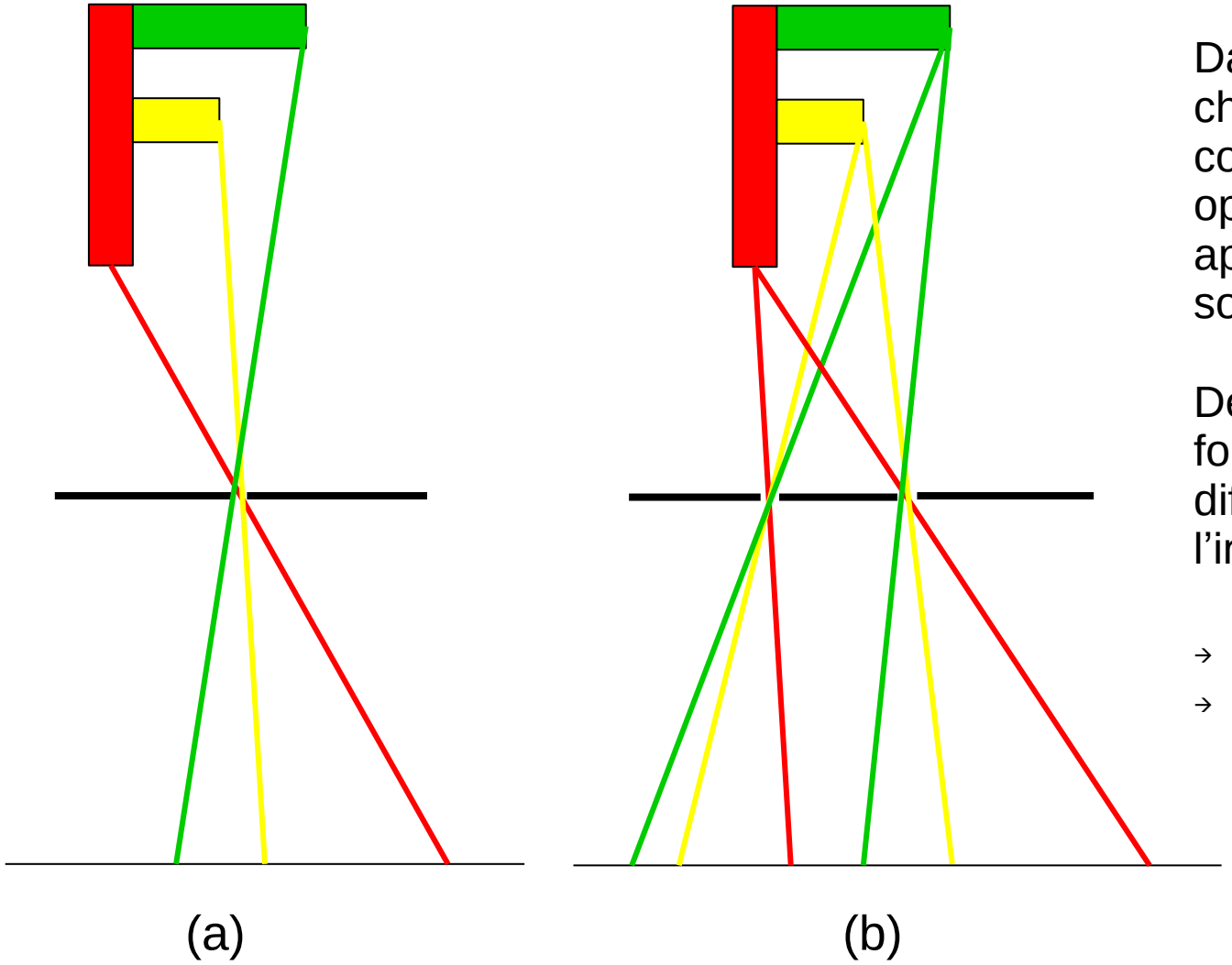
Les approches présentées dans cette partie sont toutes fondées sur l'utilisation d'une ouverture non ponctuelle associée à une lentille, en exploitant l'information de focus et de flou :

- Caméra plénoptique
- Profondeur par le focus
- Ouverture codée

Caméra plénoptique 3d Raytrix™

Camera plénoptique plein champ Lytro™

APPROCHES PASSIVES : STENOPEES…



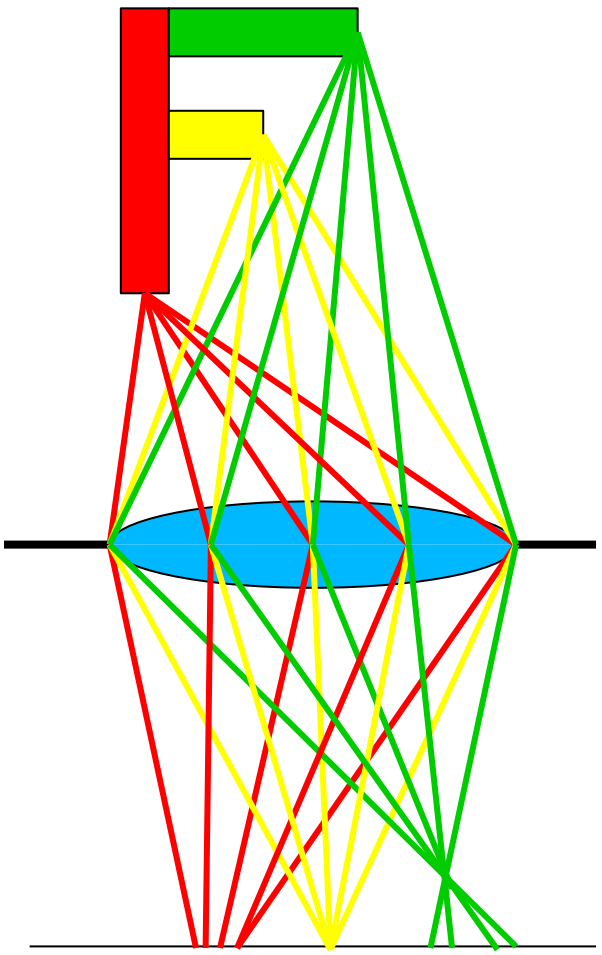
(a) (b)

Dans une caméra sténopée (a), chaque point de l'image correspond à un unique chemin optique. Tous les points apparaissent nets quelles que soient leurs profondeurs.

Deux points de vue distincts (b), font apparaître des images différentes, dont on peut extraire l'information de profondeur.

- *Stéréovision*
- *Structure from Motion*

APPROCHES PASSIVES : …*VS* LENTILLES



Dans une caméra avec lentille, chaque point de la scène illumine le plan focal selon une multitude de chemins optiques, correspondant au faisceau de droites formé par le cône ayant pour base l'ouverture du diaphragme.

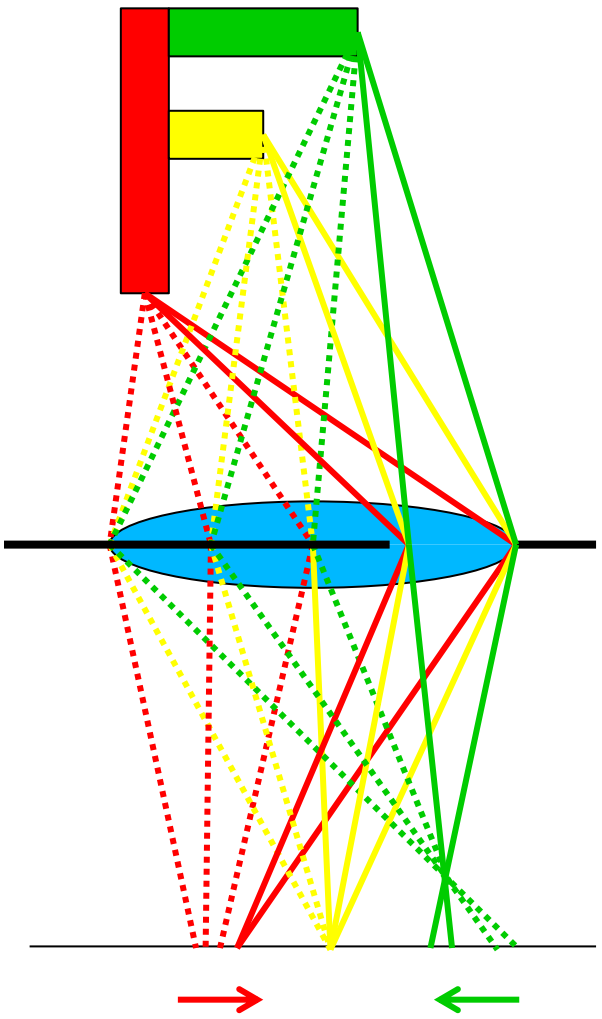
Chaque chemin correspond à une portion infinitésimale de l'ouverture, à travers laquelle on voit la scène sous un angle particulier.

Chaque portion infinitésimale forme donc une image de type sténopée, et l'image formée par la lentille correspond à la somme des différentes « images sténopées ».

Si le point est dans le plan conjugué du plan focal (plan de netteté), les différents chemins convergent sur l'image, et le point apparaît net, sinon il apparaît plus ou moins flou en fonction de la distance au plan de netteté.

→ *Depth from (de)focus*

APPROCHES PASSIVES : LENTILLE ET OUVERTURES



En utilisant une ouverture excentrique (figure), on sélectionne un sous-ensemble des chemins optiques, réduisant à la fois le flou et la quantité de lumière reçue.

Les points dans le plan de netteté (chemins jaunes) conservent leur localisation dans le plan image.

Les points plus proches (chemins rouges) sont déviés dans la direction de l'ouverture.

Les points plus lointains (chemins verts) sont déviés dans la direction opposée.

→ *Ouverture codée :*
Exploiter la géométrie de l'ouverture de façon à interpréter plus facilement le flou (≈ réponse impulsionnelle de l'ouverture).

→ *Caméra plénoptique :*
Séparer physiquement les différents chemins optiques en sous-faisceaux focalisés sur des parties distinctes du capteur.

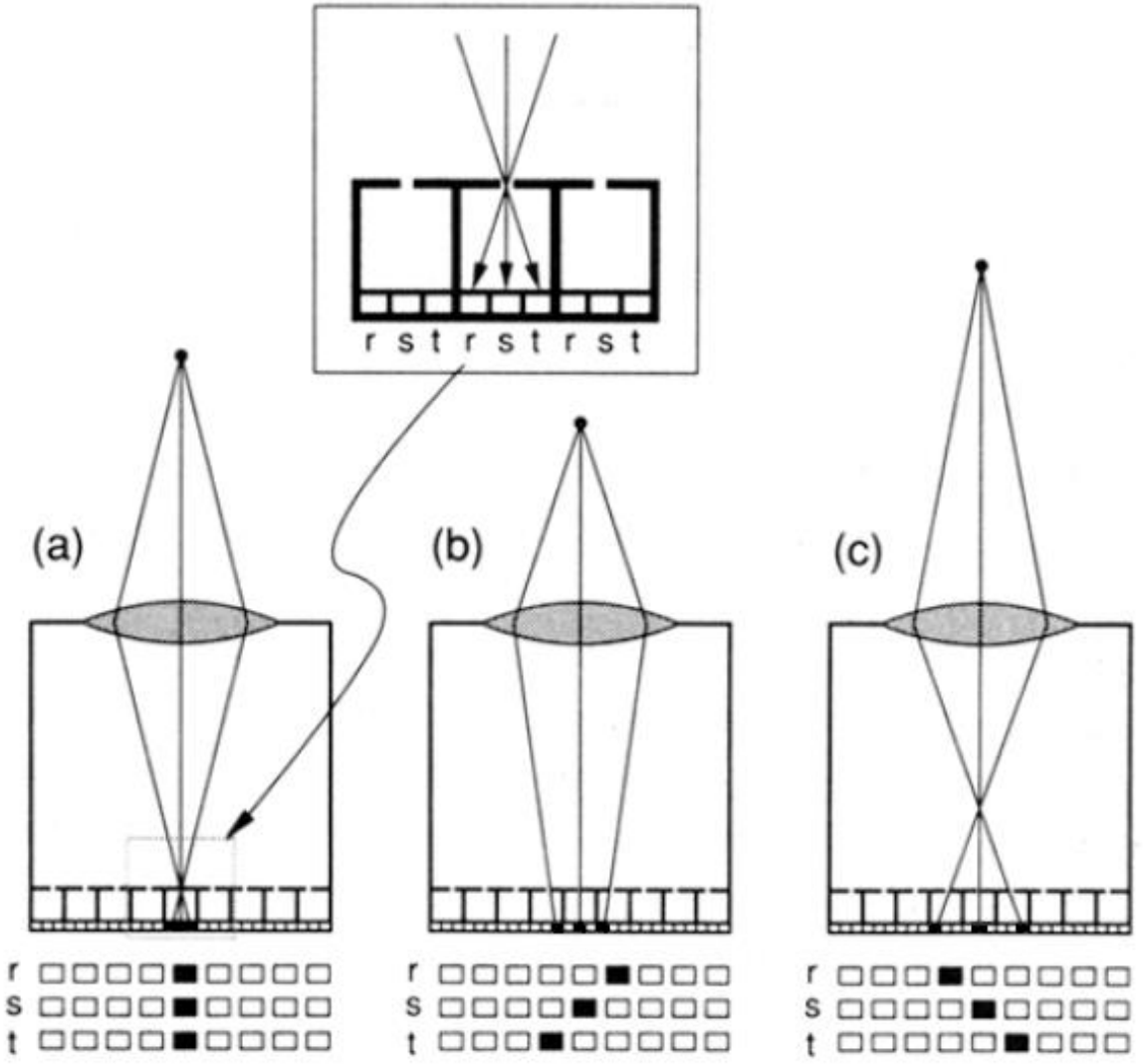
3D PASSIF : CAMERA PLENOPTIQUE

Dans une caméra plénoptique, les chemins optiques sont séparés en sous-faisceaux qui sont focalisés sur différentes parties du capteur
(Figure : mini-sténopées, mais aussi réseau lenticulaire 1d, ou encore réseau de micro-lentilles 2d).

L'information acquise se compose d'une macro-image composée d'hyper-pixels ou micro-images.
(Ex figure :

- Macro-image de 1x9 hyper-pixels.
- Hyper-pixel de taille 1x3.)

L'image plénoptique capture donc une information 4d : $I(x,y,\xi,\varsigma)$, où (x,y) correspond à la direction d'un point illuminant l'ouverture (cône de lumière), et (ξ,ς) un point de vue particulier de ce point à travers l'ouverture.

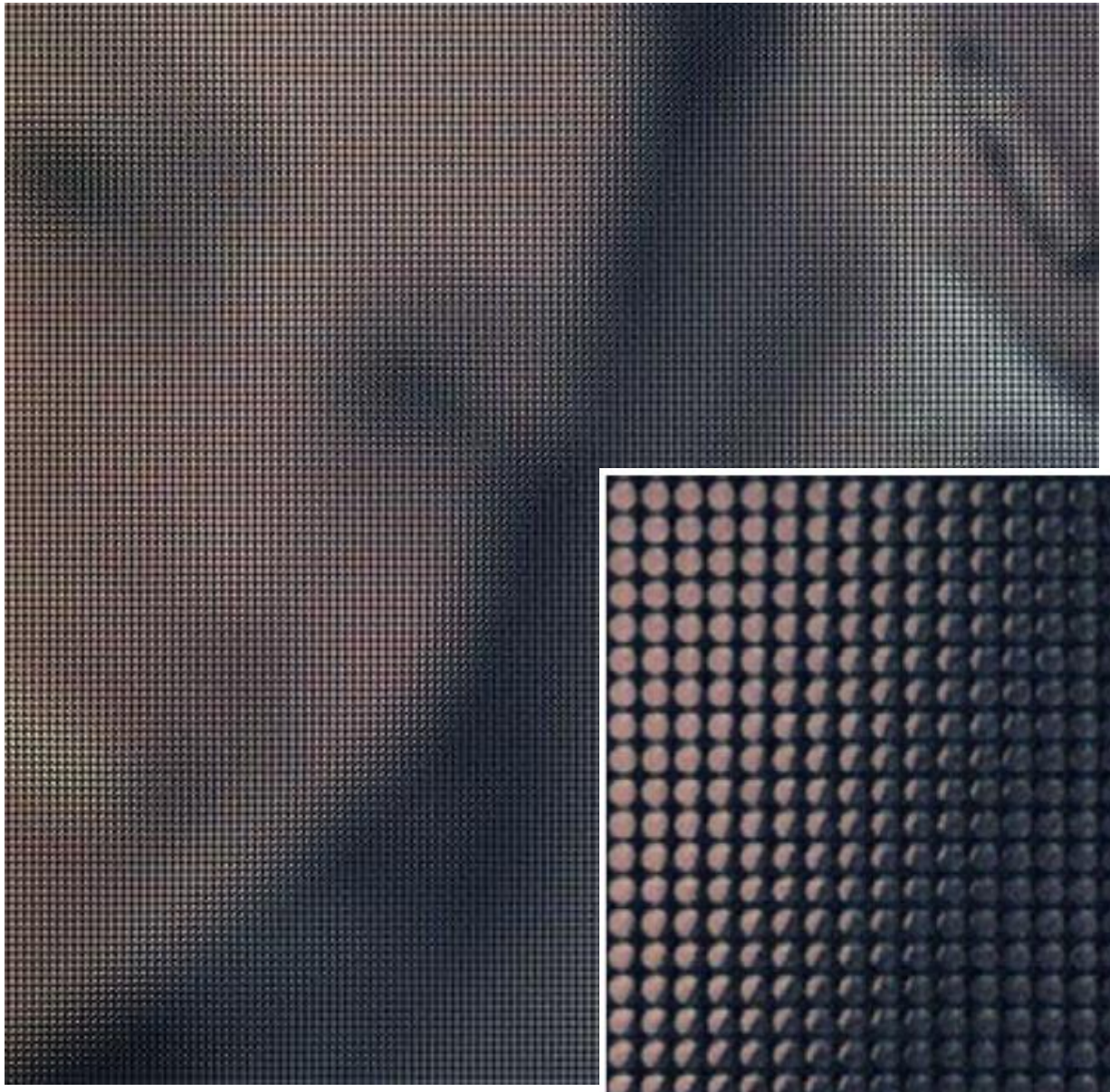


[Adelson 1992]

CAMERA PLENOPTIQUE : MACRO-IMAGE

[Ng 2005]

CAMERA PLENOPTIQUE : MICRO-IMAGES

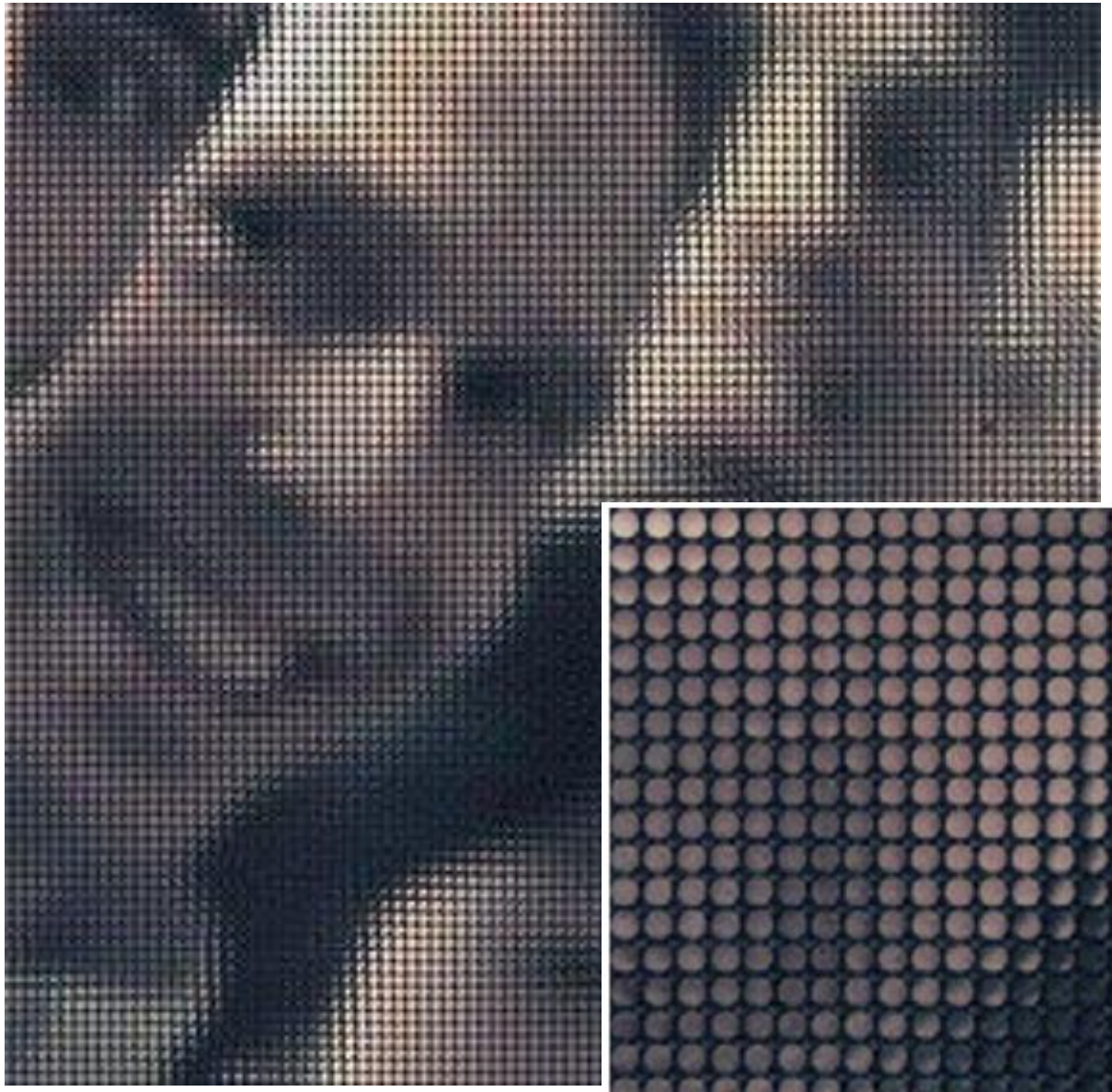


(a)

Les micro-images (hyper-pixels) des points plus proches que le plan de netteté *présentent un contraste dans le même sens* que la macro-image.

[Ng 2005]

CAMERA PLENOPTIQUE : MICRO-IMAGES

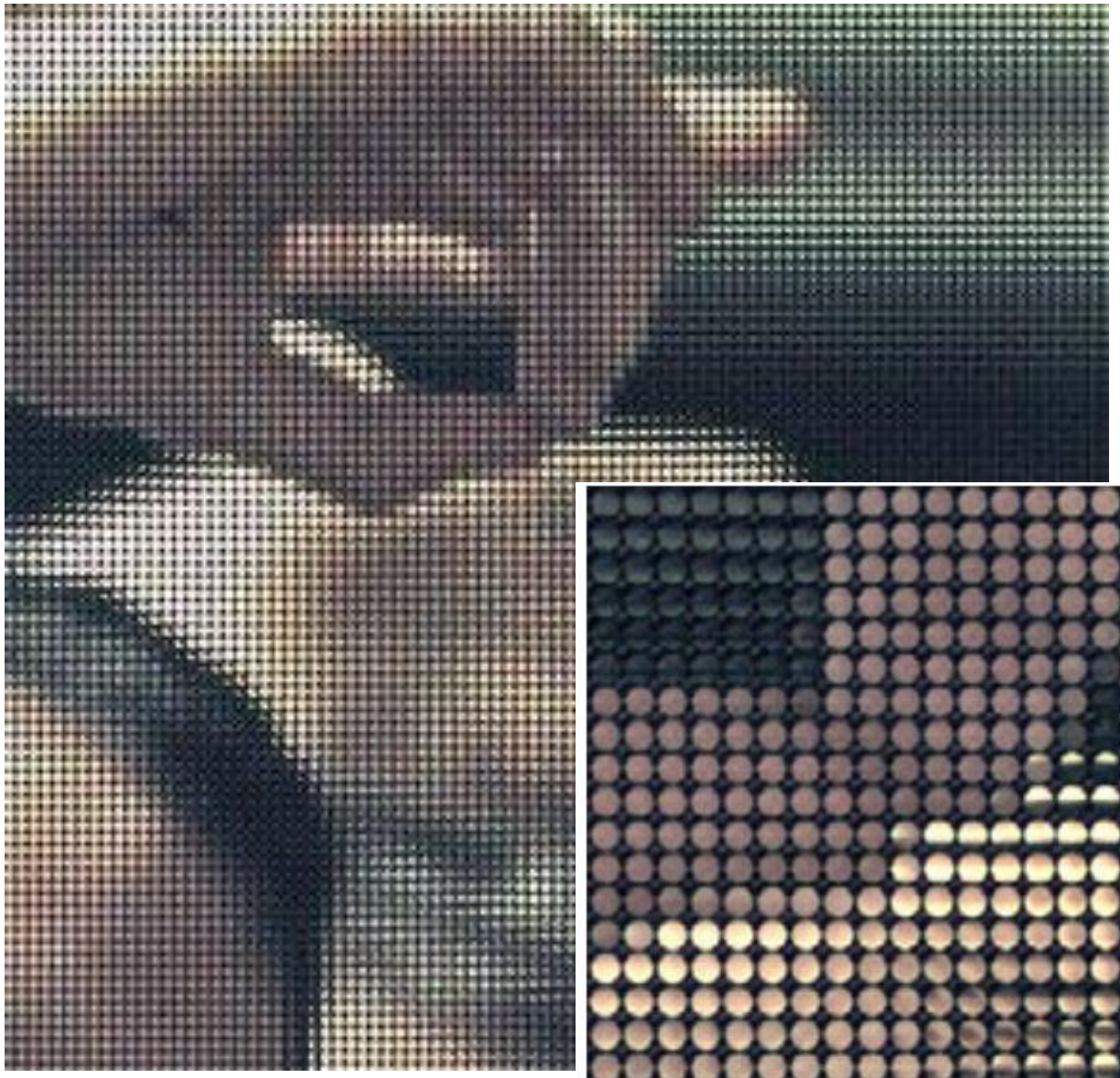


(b)

Les micro-images (hyper-pixels) des points plus lointains que le plan de netteté *présentent un contraste en sens* inverse de la macro-image.

[Ng 2005]

CAMERA PLENOPTIQUE : MICRO-IMAGES

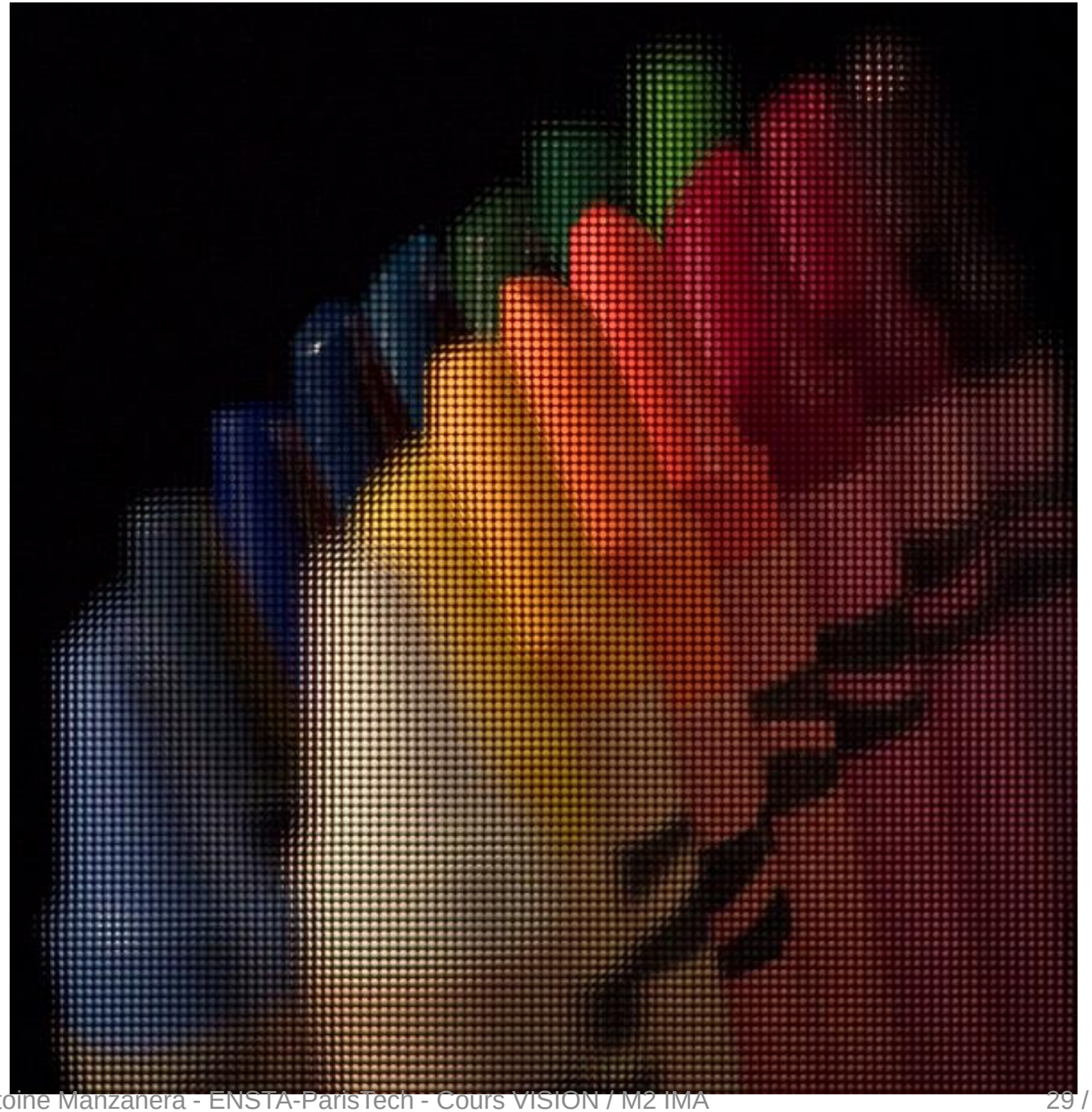


(c)

Les micro-images (hyper-pixels) des points dans le plan de netteté *forment des régions homogènes.*

[Ng 2005]

PLENOPTIQUE : MACRO-IMAGE

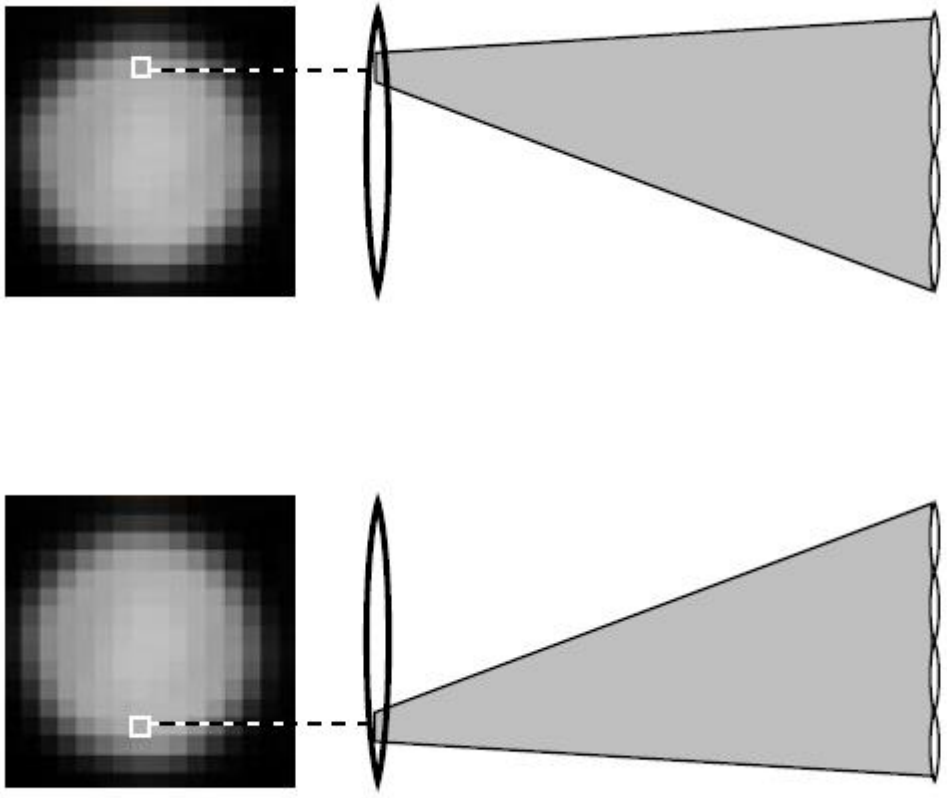


[Ng 2005]

PLENOPTIQUE : MACRO-IMAGE ET MACRO-IMAGES DUALES

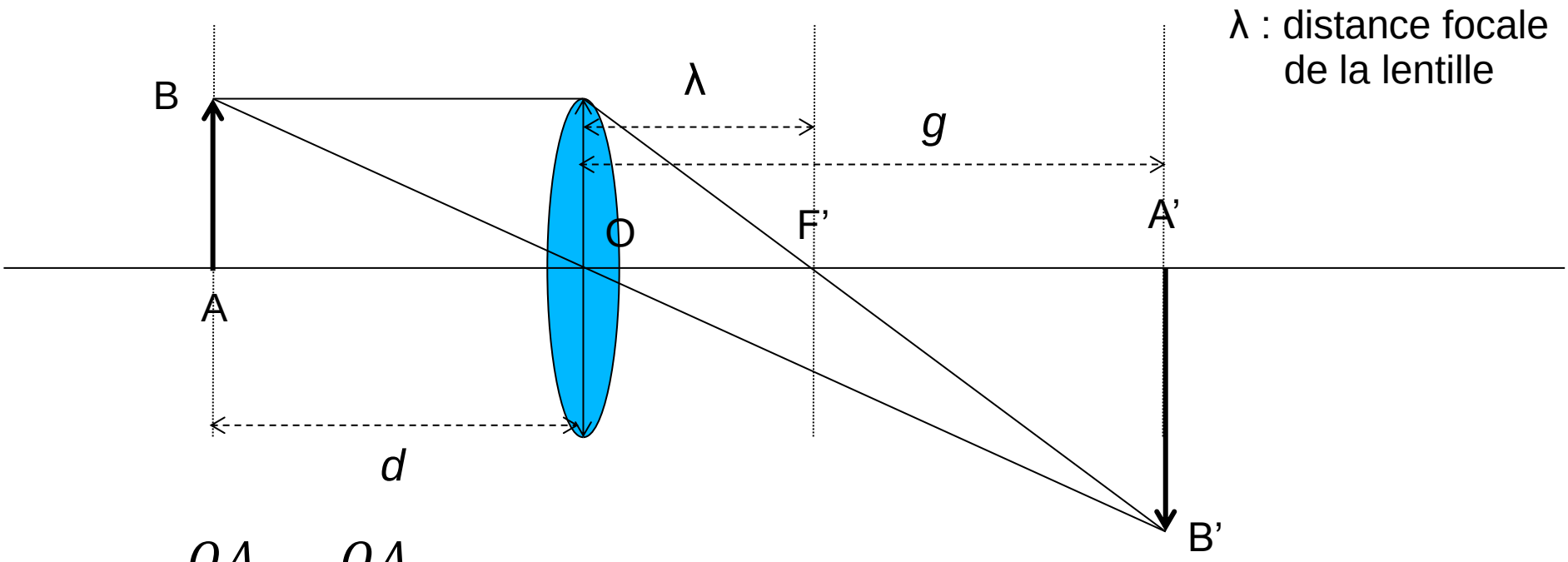
Les macro-images duales sont formées en recomposant *n* x *n* images sous-résolues de taille *m* x *m* à partir des pixels homologues de chaque micro-image, ou *n* x *n* est le nombre de micro-images (résolution de la macro-image), et *m* x *m* la résolution de la micro-image.

Les macro-images duales correspondent à une partition de l'ouverture du diaphragme en points de vue distincts et présentent donc des différences de parallaxe dont on peut déduire l'information de profondeur par appariement (*single-lens stereo*).



[Ng 2005]

GEOMETRIE DE LA LENTILLE MINCE CONVERGENTE



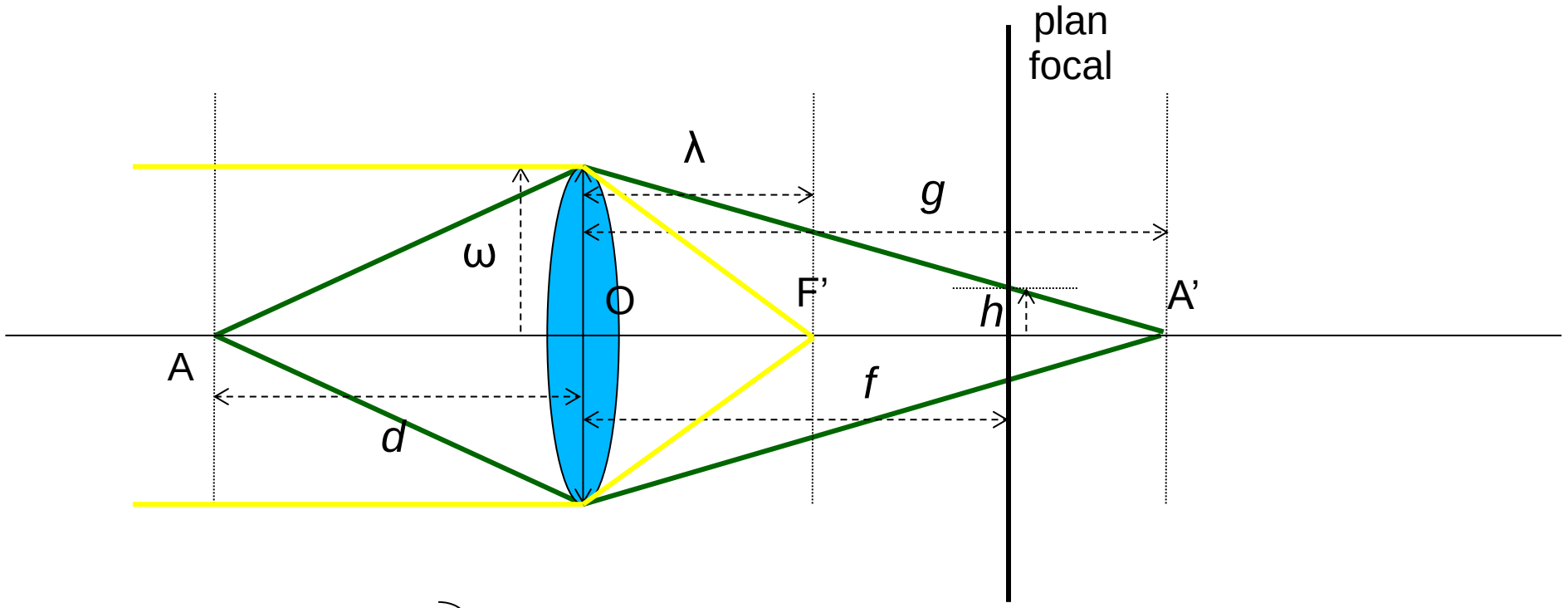
$$\frac{OA}{OF'} = \frac{OA}{OA'} + 1$$

et donc :

$$\boxed{\frac{1}{\lambda} = \frac{1}{d} + \frac{1}{g}}$$

Équation de la lentille mince.

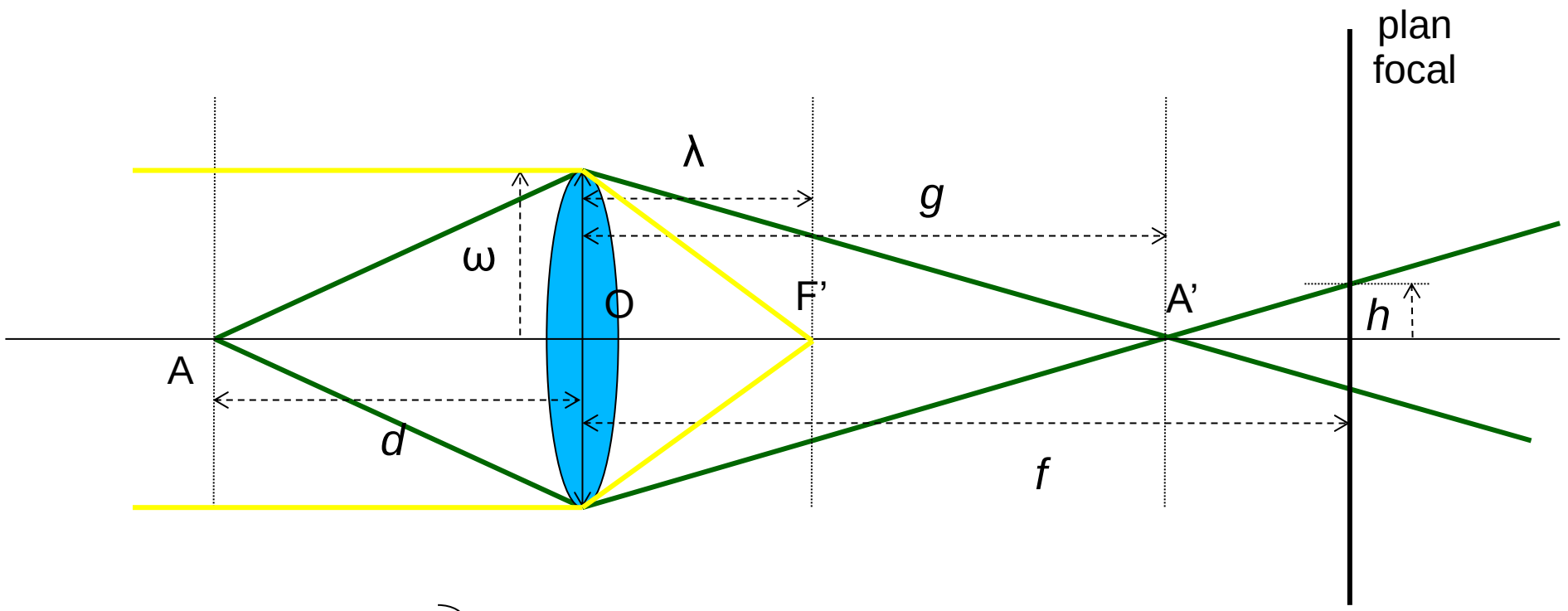
RELATION FOCUS / DISTANCE : FOCALE COURTE



$$\left.\begin{aligned} \frac{f}{g} &= 1 - \frac{h}{\omega} \\ \frac{1}{\lambda} &= \frac{1}{d} + \frac{1}{g} \end{aligned}\right\} \quad \frac{1}{d} = \left(\frac{1}{\lambda} - \frac{1}{f}\right) + \frac{h}{f\omega}$$

λ : focale lentille
ω : ouverture
f : focale image
h : largeur défocus

RELATION FOCUS / DISTANCE : FOCALE LONGUE



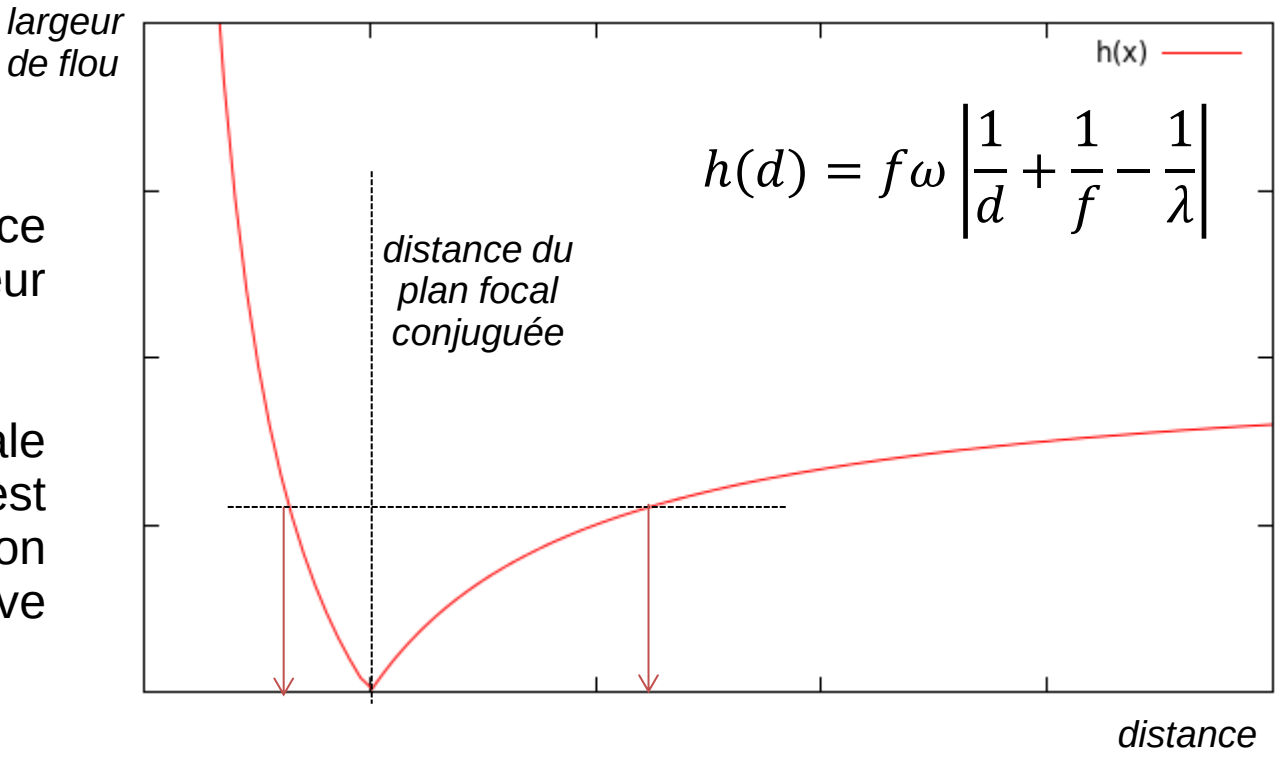
$$\left.\begin{aligned} \frac{f}{g} &= 1 + \frac{h}{\omega} \\ \frac{1}{\lambda} &= \frac{1}{d} + \frac{1}{g} \end{aligned}\right\} \quad \frac{1}{d} = \left(\frac{1}{\lambda} - \frac{1}{f}\right) - \frac{h}{f\omega}$$

λ : focale lentille
ω : ouverture
f : focale image
h : largeur défocus

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS

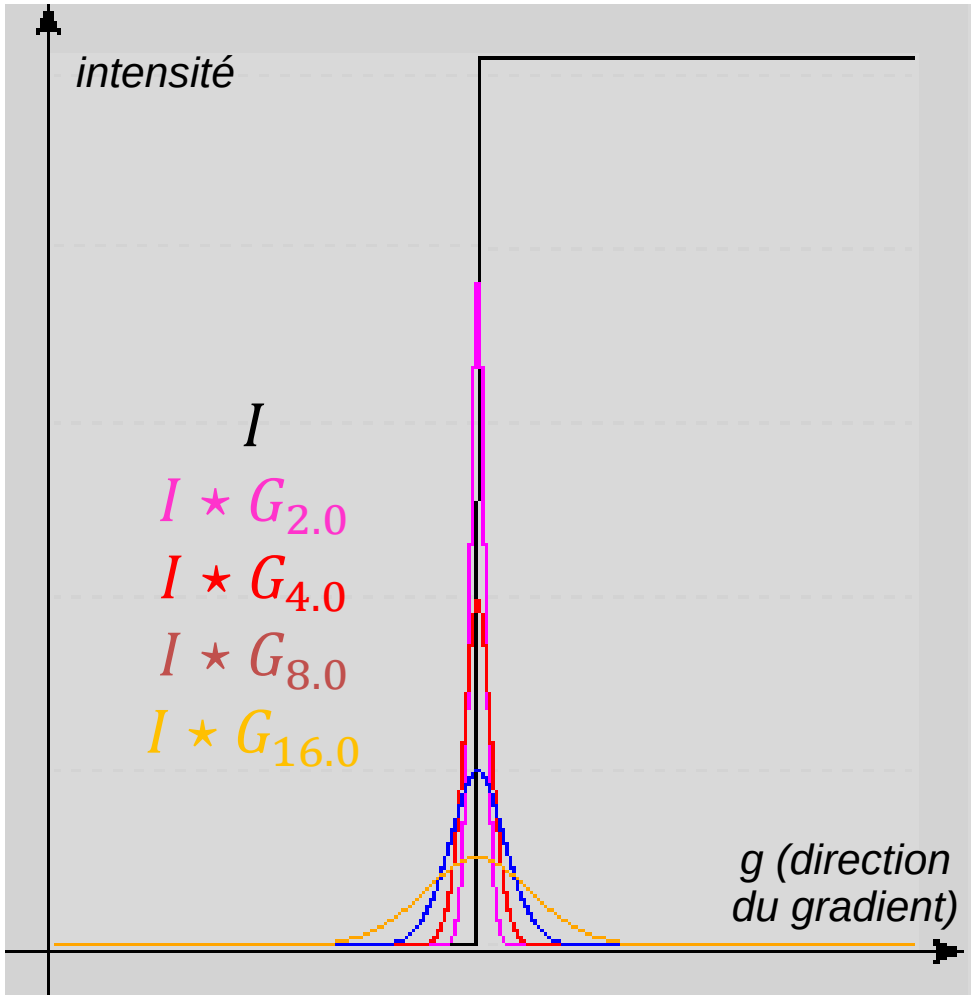
L'estimation de la distance revient donc à estimer la largeur de flou dans l'image.

Sans hypothèse sur la focale image, une mesure unique est toujours ambiguë, la fonction *h(d)* n'étant pas injective (Figure).



Pour une mesure directe de la largeur de flou par traitement d'image, une hypothèse sur la structure de l'image nette est nécessaire : impulsion, contour type échelon, de façon à pouvoir prédire l'effet du flou sur cette structure.

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS



Si l'on modélise le flou par une convolution gaussienne 2d dont l'écart-type est fonction de *h*, on peut déduire *h* de l'effet produit sur une structure contour de type « échelon », en mesurant la valeur du maximum local du gradient dans la direction orthogonale à l'échelon.

Ces structures correspondent à la définition classique des contours, i.e. les passages par zéro de la dérivée seconde dans la direction du gradient *g*:

$$C_I = \left\{ x; \frac{\partial^2 I}{\partial g^2}(x) = 0 \right\}$$

Question : comment justifier l'utilisation d'un modèle de flou gaussien alors que l'optique géométrique conduit à une fonction porte ?

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS

$I(x) = (I^H(x), I^S(x), I^V(x))$

$\frac{\partial I^V}{\partial g}(x)$ (module du gradient)

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS

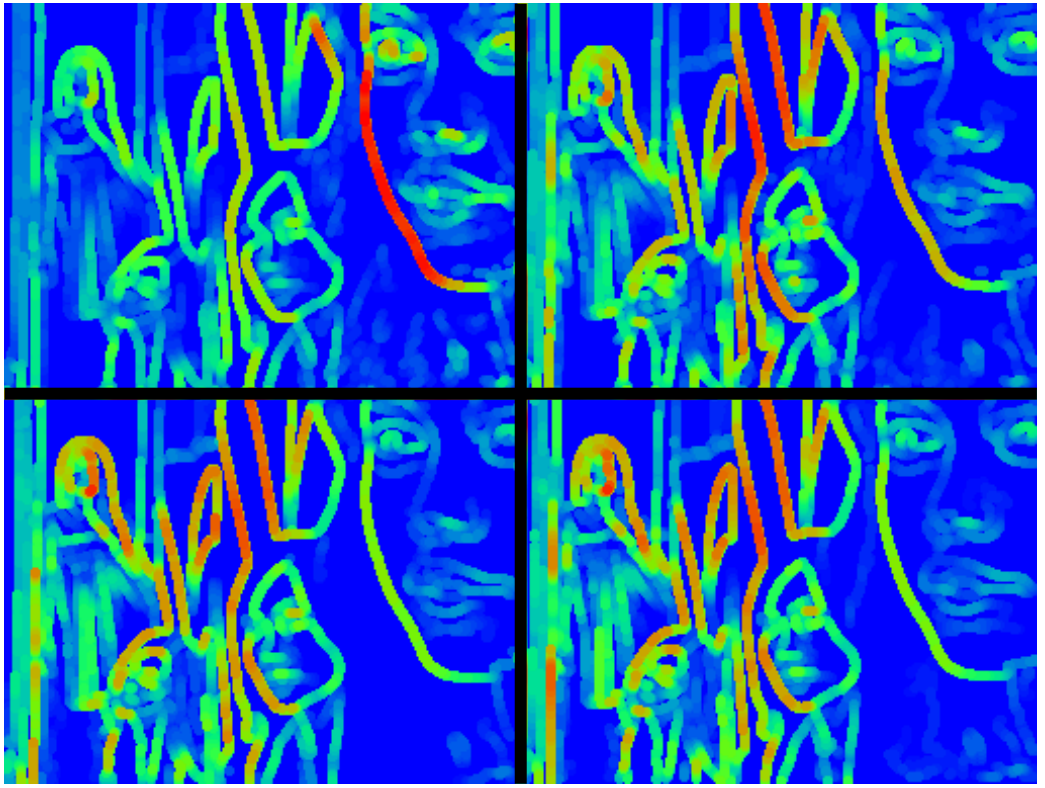
$I(x)$

$C_I = \left\{x; \frac{\partial^2 I^V}{\partial g^2}(x) = 0\right\}$ (contours)

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS

La mesure du module du gradient sur les contours permet une estimation de la largeur de flou *h*, mais cette estimation reste ambiguë vis-à-vis de la position par rapport au plan de netteté.

$I(x)$



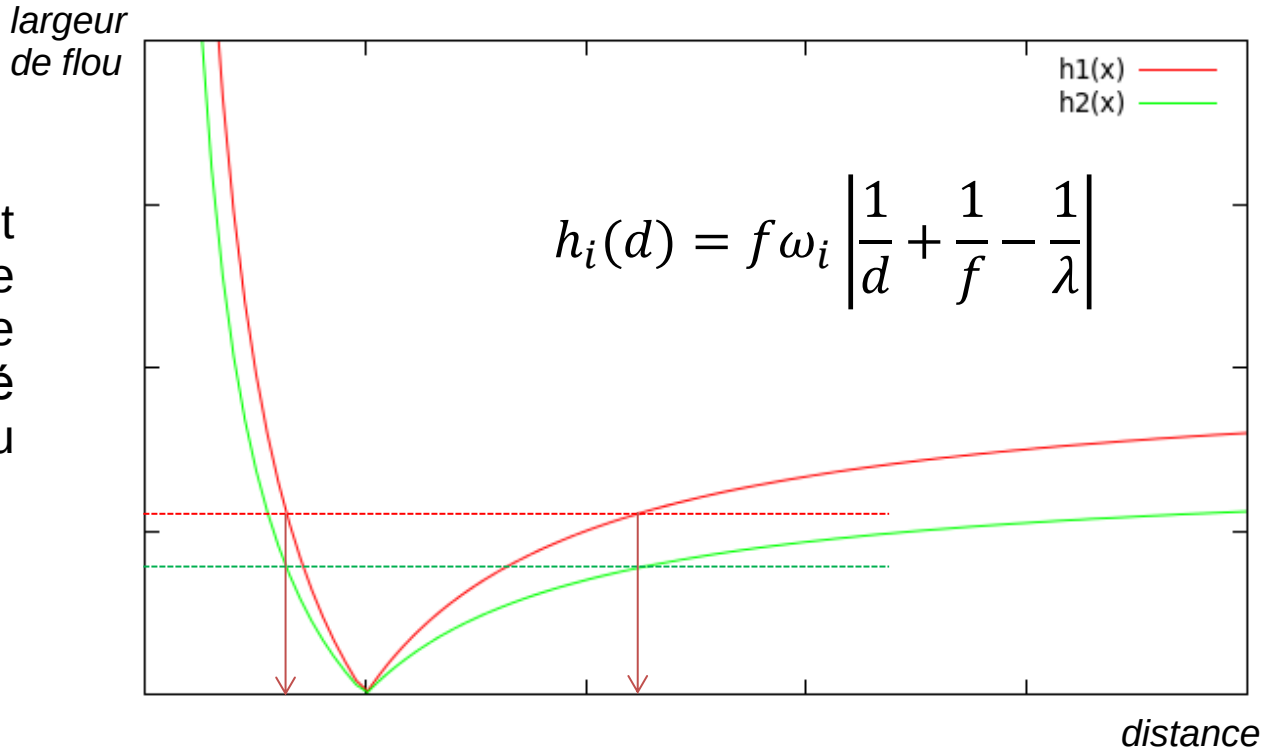
Mesure du flou le long des contours

[Pentland 1987]

Idée : répéter la mesure en faisant varier l'ouverture ω et/ou la focale image f ?

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS

Le flou dépendant linéairement de l'ouverture, l'utilisation de différentes ouvertures seule ne permet pas de lever l'ambiguïté de la distance par rapport au plan de netteté :



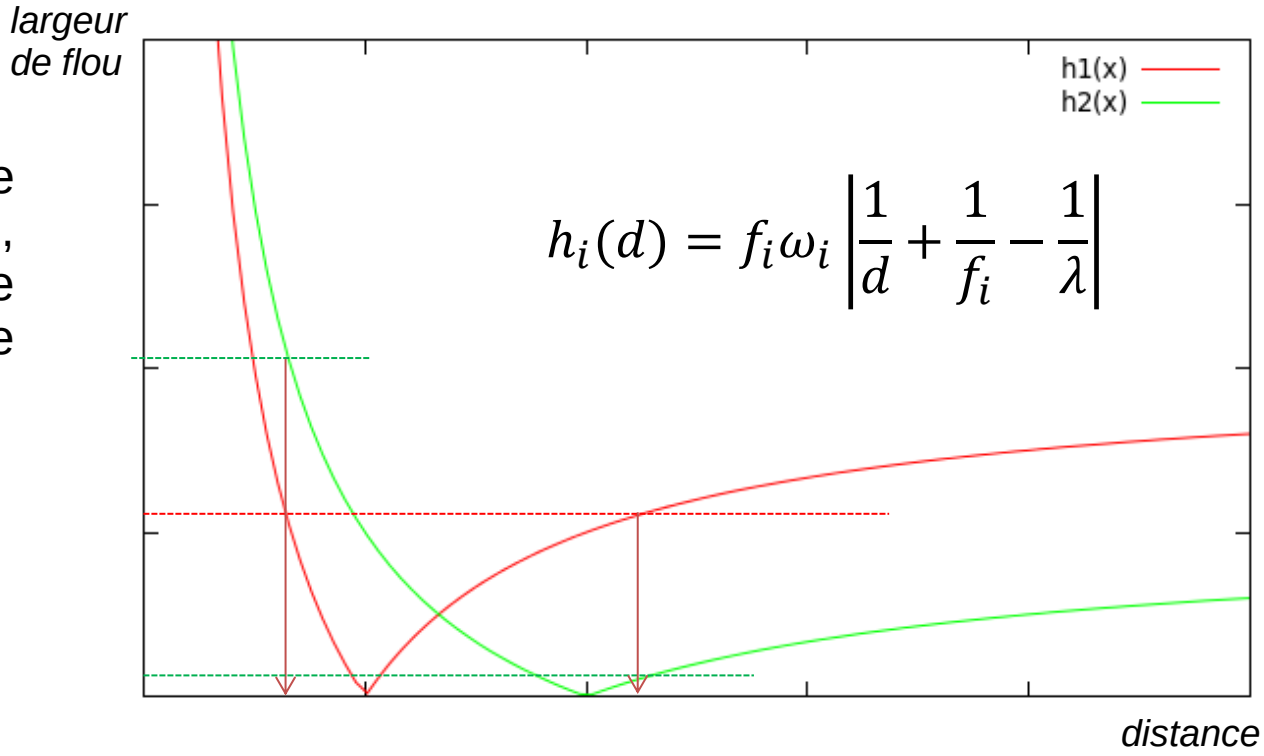
Focale constante, ouverture variable

3D PASSIF : PROFONDEUR PAR LE (DE)FOCUS

En revanche, l'utilisation de plusieurs couples (ouverture, focale image) permet de déduire la distance du flou de façon absolue.

(Figure : produit $f_i \omega_i$ constant)

[Pentland 1987]



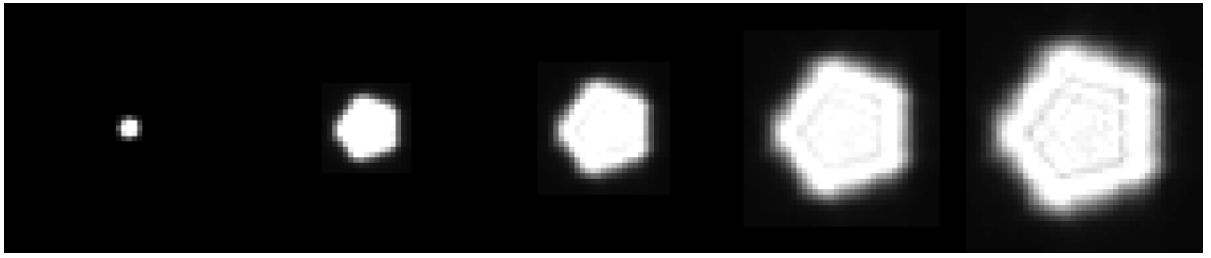
Ouverture constante, focale variable

MODELE DE FLOU *VS* ETALONNAGE D'OUVERTURE

Le noyau gaussien peut être un meilleur modèle de flou que la fonction porte à cause de la combinaison de plusieurs phénomènes : diffraction, aberrations chromatiques, discrétisation, qui se traduit par la composition de plusieurs convolutions.

Cependant, une bonne alternative à l'utilisation d'un modèle de flou est de réaliser un étalonnage de l'ouverture de la caméra en enregistrant les différentes images formées par un point pour différentes distances de focalisation (réponses impulsionnelles des noyaux de convolution).

[Levin 2007]



Diaphragme traditionnel à 5 lames et famille $\{g_d\}_{d \in D}$ *de noyaux étalonnée.*

L'estimation de la bonne distance revient alors à trouver le noyau g_d qui correspond le mieux à ce qu'on observe localement.

L'estimation « directe » ne pouvant se faire que sur les contours, on utilise plutôt l'estimation indirecte par déconvolution…

ESTIMATION DE FLOU PAR DECONVOLUTION

I l'image observée

$\{g_d\}_{d\in D}$ la famille des noyaux de convolution étalonnés, indexée par la distance

J_d la déconvolution de I par g_d

L'erreur de reconstruction $\varepsilon_d(x)$ au pixel x et à la distance d est définie par :

$$\varepsilon_d(x) = \sum_{y\in W_x} \|I - J_d \star g_d\|^2$$

où W_x est un voisinage de x.

L'estimation de la distance est réalisée par :

$$d_{opt}(x) = \arg\min_{d\in D} \varepsilon_d(x)$$

DECONVOLUTION : FILTRE INVERSE ET FILTRE DE WIENER

Le problème se ramène donc à une déconvolution d'image (restauration), connaissant le noyau de convolution à l'origine du flou.

Rappel du cours de Restauration :

$$F = I \star g_d \xrightarrow{\text{transformée de Fourier}} \tilde{F} = \tilde{I} \times \widetilde{g_d} \xrightarrow{\text{filtrage inverse}} \widetilde{J_d} = \frac{\widetilde{F}}{\widetilde{g_d}} \xrightarrow{\text{transformée de Fourier inverse}} J_d$$

Inutilisable à cause des zéros de $\widetilde{g_d}$ et du bruit additif !!!

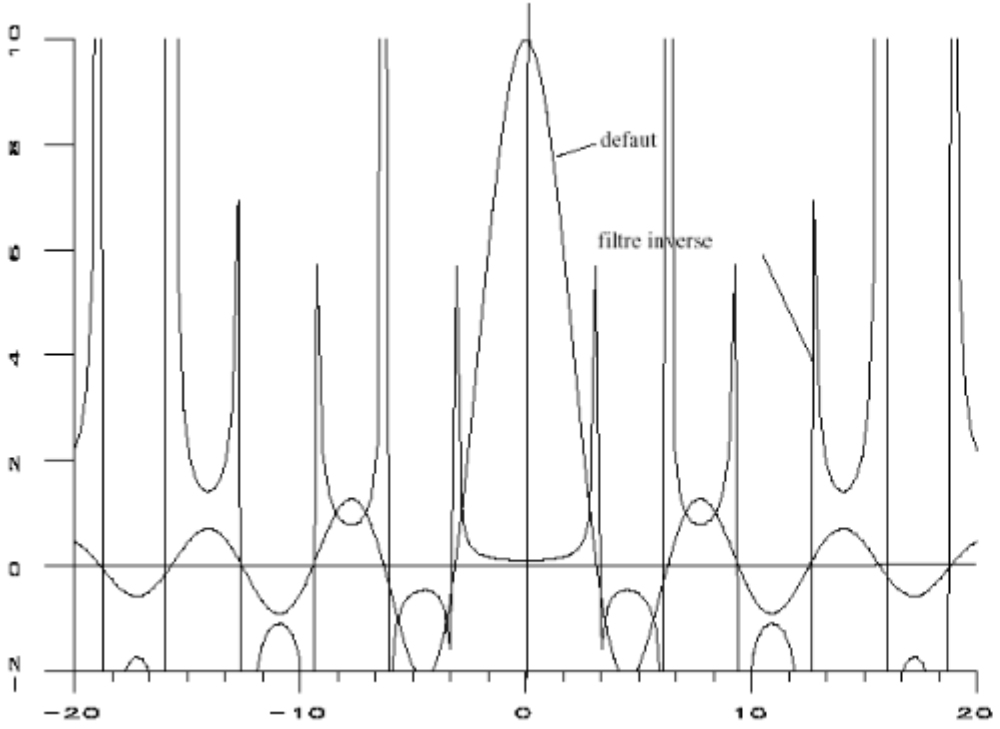
$$F = I \star g_d + b \xrightarrow{\text{transformée de Fourier}} \tilde{F} = \tilde{I} \times \widetilde{g_d} + \tilde{b} \xrightarrow{\text{filtrage de Wiener}} \widetilde{J_d} = \frac{\widetilde{g_d}' \times \tilde{F}}{\widetilde{g_d}\widetilde{g_d}' + \alpha} \xrightarrow{\text{transformée de Fourier inverse}} J_d$$

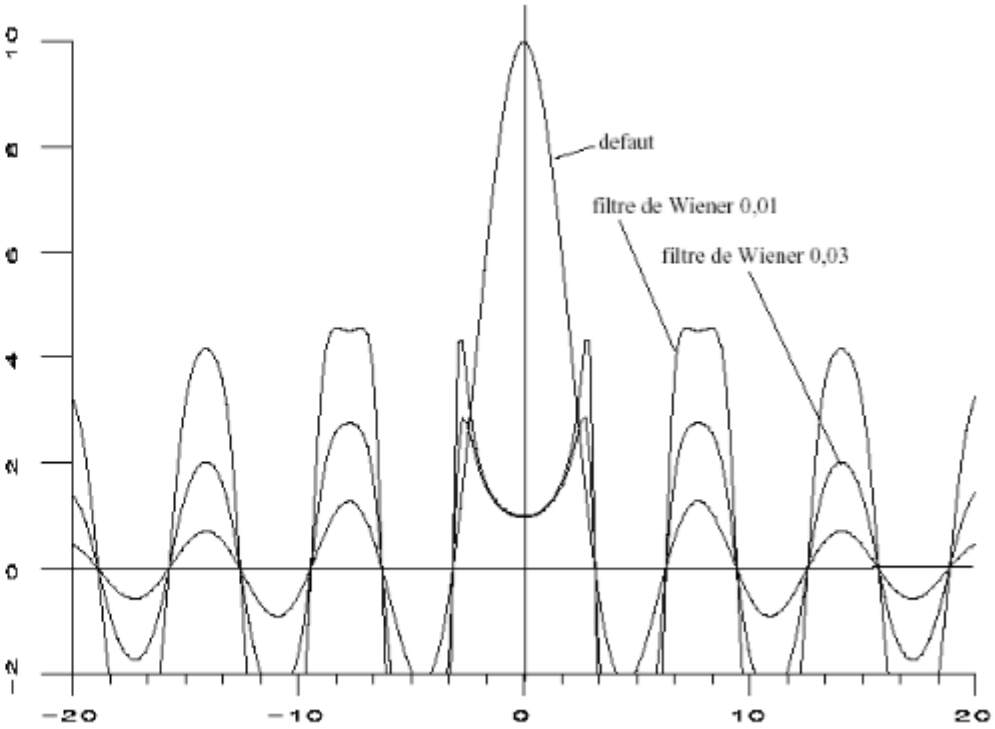
$$\alpha \approx \frac{\left\langle \left|\tilde{b}(u)\right|^2 \right\rangle}{\left\langle \left|\tilde{I}(u)\right|^2 \right\rangle}$$

α est un terme de régularisation, qui dépend de la puissance relative du bruit *b* par rapport au signal image *I*. Il peut être constant ou dépendre des fréquences : $\alpha(u)$. Le filtrage de Wiener réalise ainsi un compromis entre déconvolution et régularisation.

Dans tous les cas, les zéros du filtre dans le domaine fréquentiel ($\widetilde{g_d}$) conditionnent fortement l'erreur de reconstruction ε_d.

DECONVOLUTION : FILTRE INVERSE ET FILTRE DE WIENER





A gauche : un défaut de bougé à vitesse constante dans le domaine fréquentiel (sinus cardinal), et le filtre inverse correspondant.
A droite : le même défaut et les filtres de Wiener de correction pour deux valeurs différentes de α supposé constant.

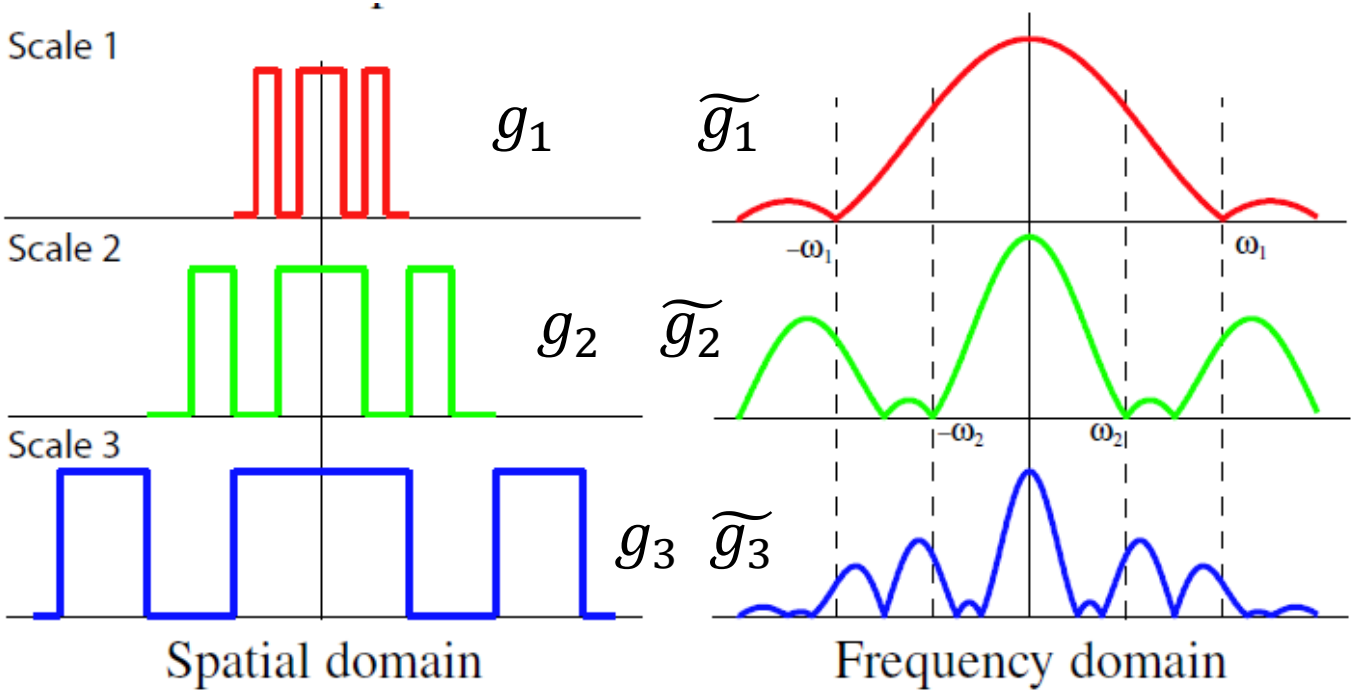
[Figure : Maître 2003]

3D PASSIF : OUVERTURE CODEE

Dans les techniques de déconvolution, ce sont les zéros du filtre dans le domaine fréquentiel qui contribuent majoritairement aux erreurs de reconstruction.

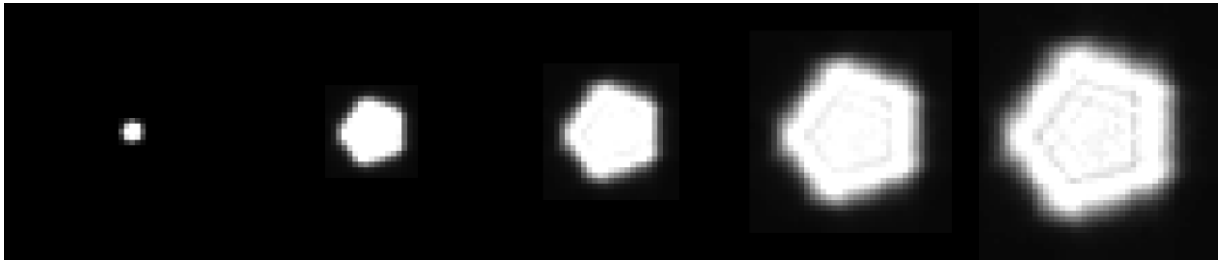
En conséquence, si les différents noyaux de convolutions candidats $\{g_d\}_{d\in D}$ ont leurs zéros qui sont placés aux mêmes endroits du domaine fréquentiel, alors il est plus difficile de distinguer leurs effets sur l'image (par déconvolution) que si leurs zéros apparaissent à des fréquences différentes.

Le principe de l'ouverture codée consiste donc à choisir une forme de diaphragme telle que les zéros des différentes versions de filtres $\{g_d\}_{d\in D}$ apparaissent, selon les différentes distances *d*, à des positions différentes du domaine fréquentiel :

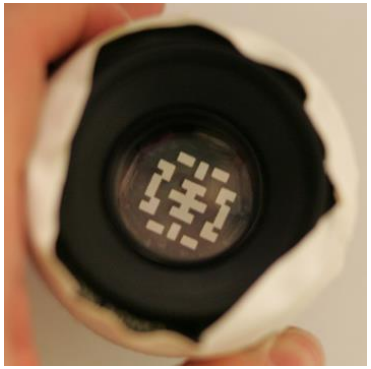


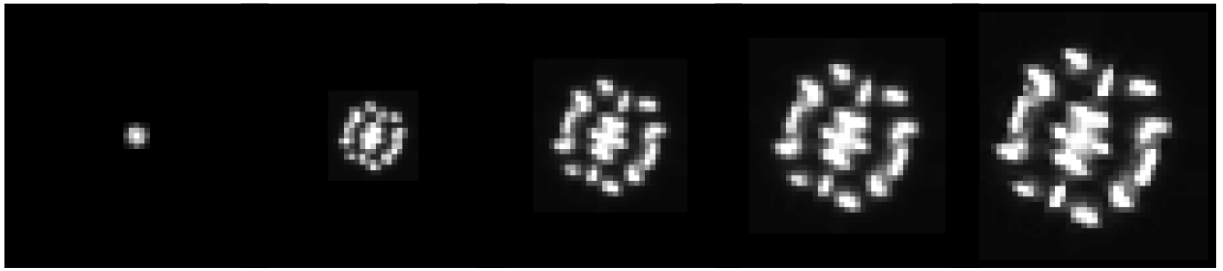
[Levin 2007]

3D PASSIF : OUVERTURE CODEE



Diaphragme traditionnel à 5 lames et famille $\{g_d\}_{d\in D}$ de noyaux.

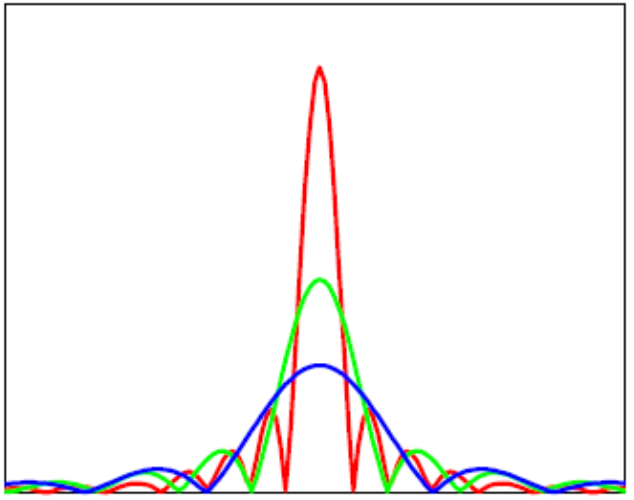




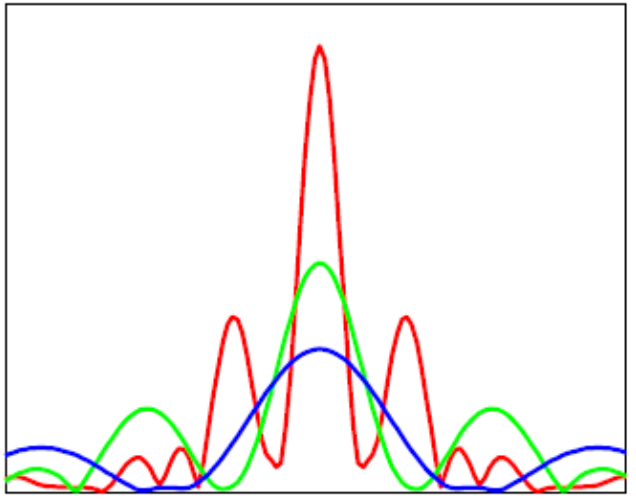
Ouverture codée et famille $\{g_d\}_{d\in D}$ de noyaux.

Comparaison des noyaux dans le domaine fréquentiel $\{\widetilde{g_d}\}_{d\in D}$ entre ouverture classique et ouverture codée (noter la position des zéros) :

[Levin 2007]



Conventional aperture



Coded aperture

3D PASSIF : OUVERTURE CODEE

Les images obtenues par déconvolution avec l'ouverture codée permettent de mieux discriminer les échelles (distances) correctes :

[Levin 2007]

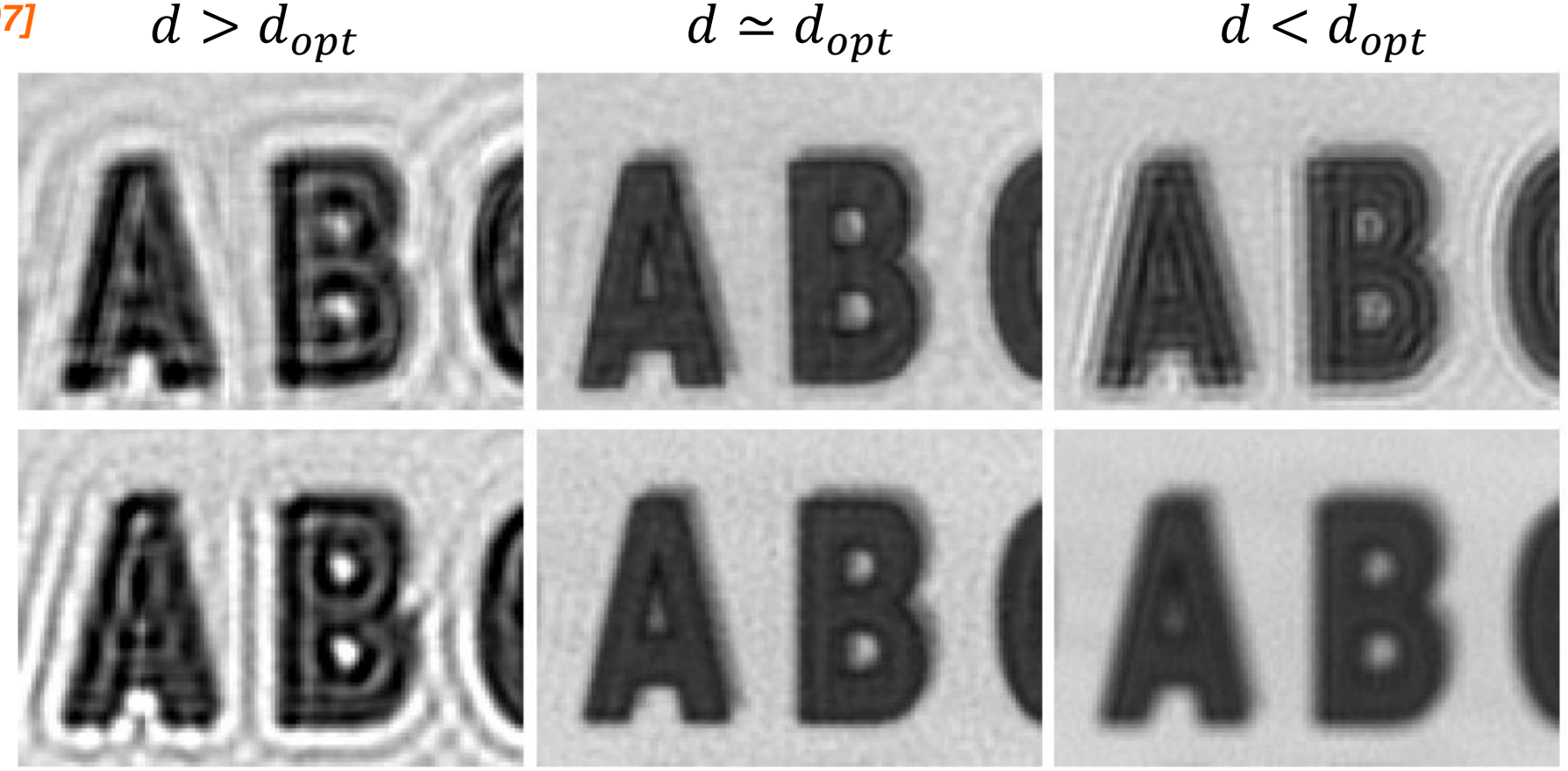
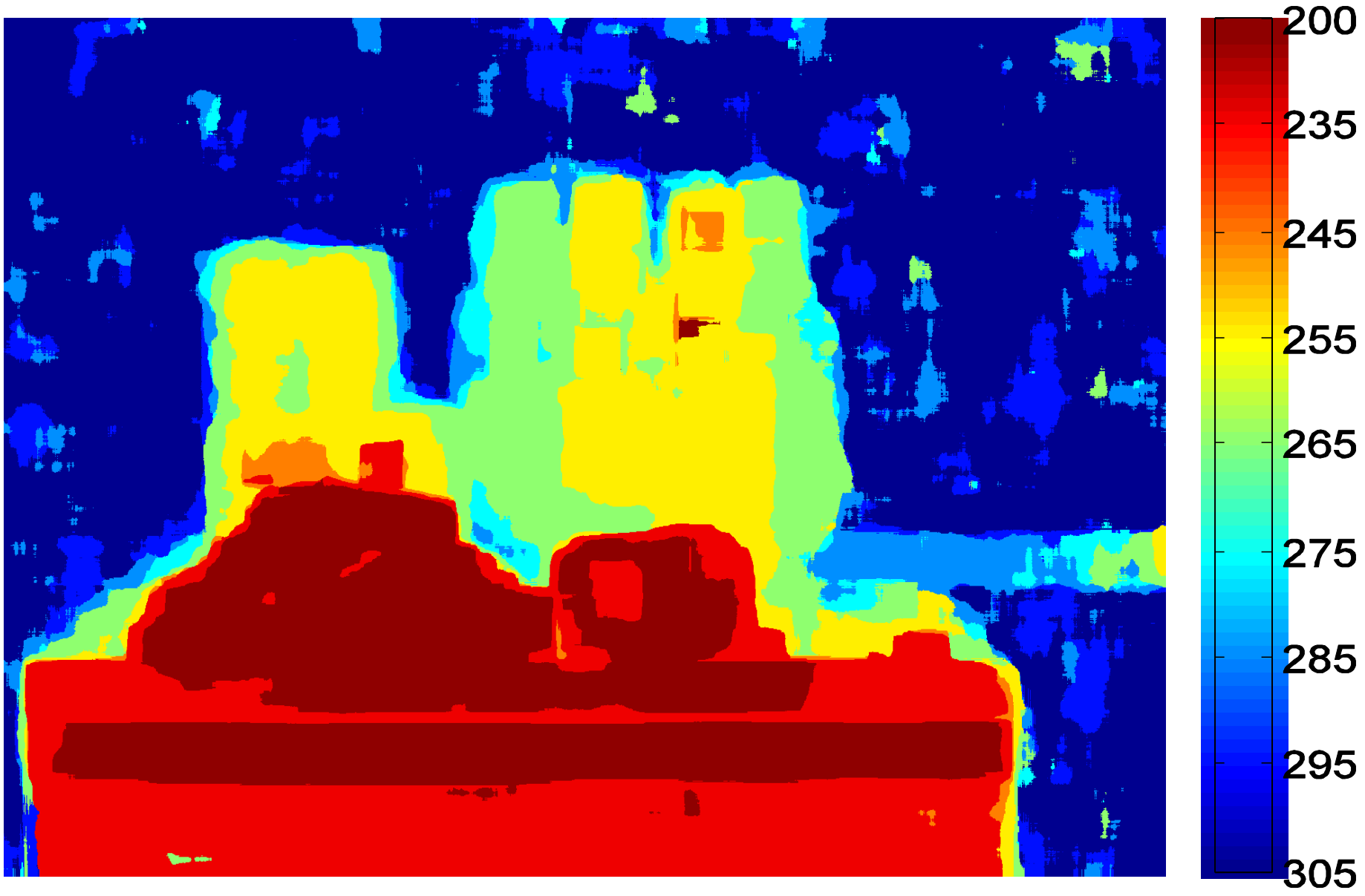


IMAGE DE DISTANCE : IMAGE TEST

[Levin 2007]

DISTANCE : RESULTAT BRUT OUVERTURE CODEE



[Levin 2007]

DISTANCE : RESULTAT APRES TRAITEMENT

[Levin 2007]

3ème Partie : TRAITEMENT D'IMAGE DANS LE PLAN FOCAL

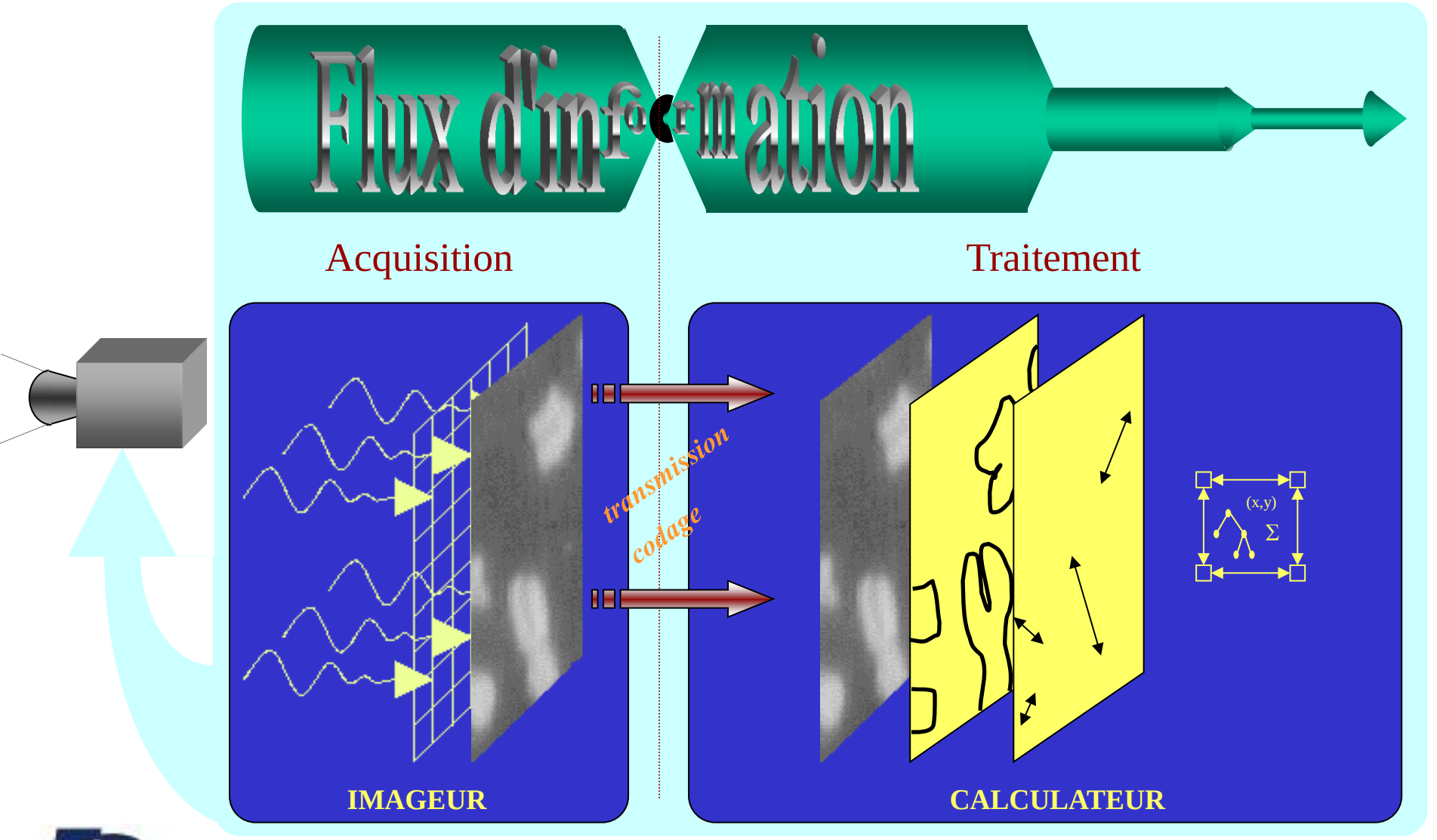
Les techniques de co-conception visent à optimiser globalement un système de vision par une démarche opportuniste qui exploite les différents éléments du système et cherche à les combiner de façon plus intime.

Les rétines artificielles s'inscrivent dans cette démarche, en cherchant à étendre l'emploi de l'électronique du capteur, de l'acquisition vers le traitement, réalisant donc l'analyse des images dans le plan focal.

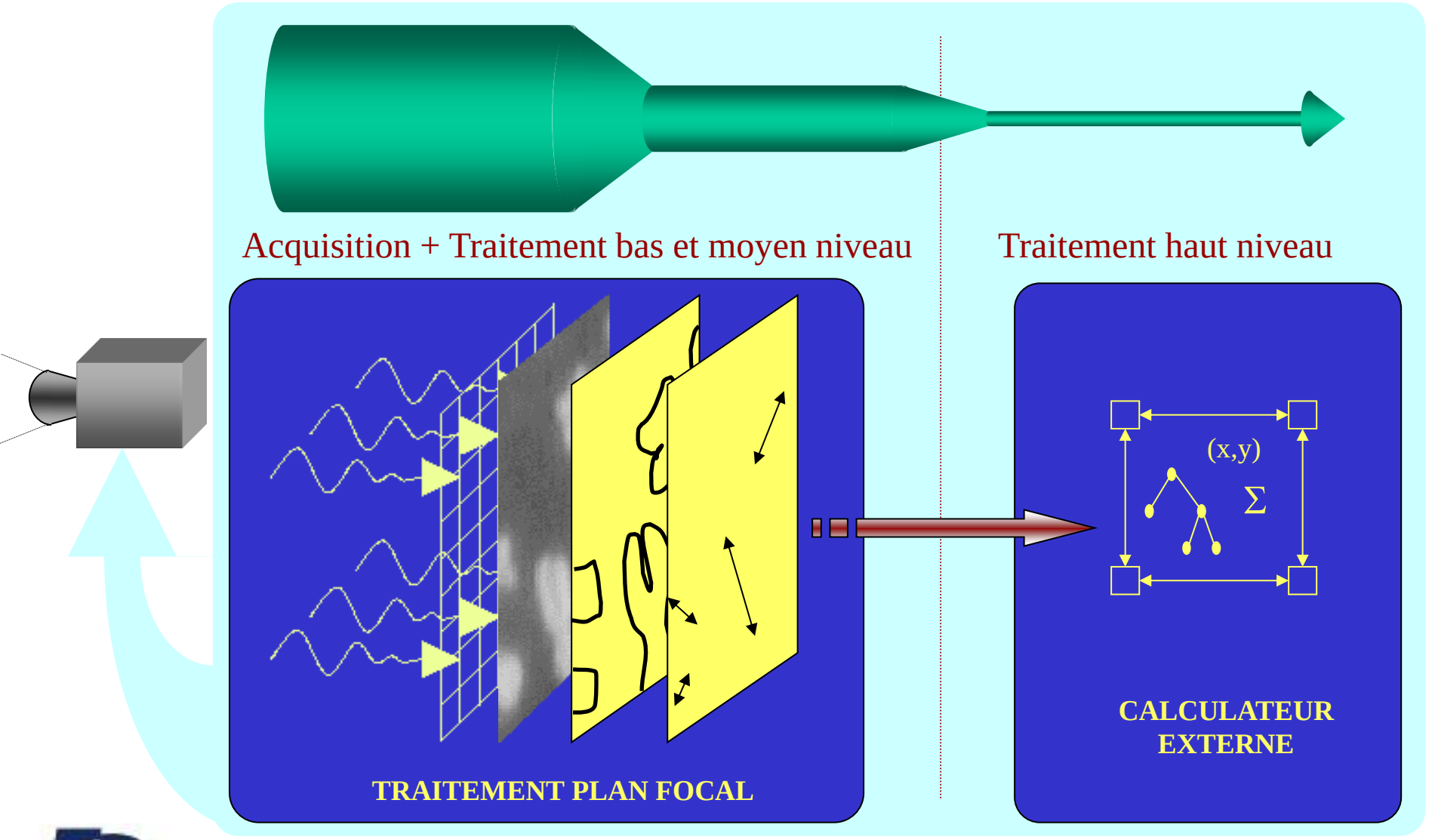
L'intérêt est de réduire au maximum le temps de calcul et/ou l'énergie consommée, grâce à :

1. La réduction drastique du débit de données transférées.
2. Le recours à un parallélisme de données massif.

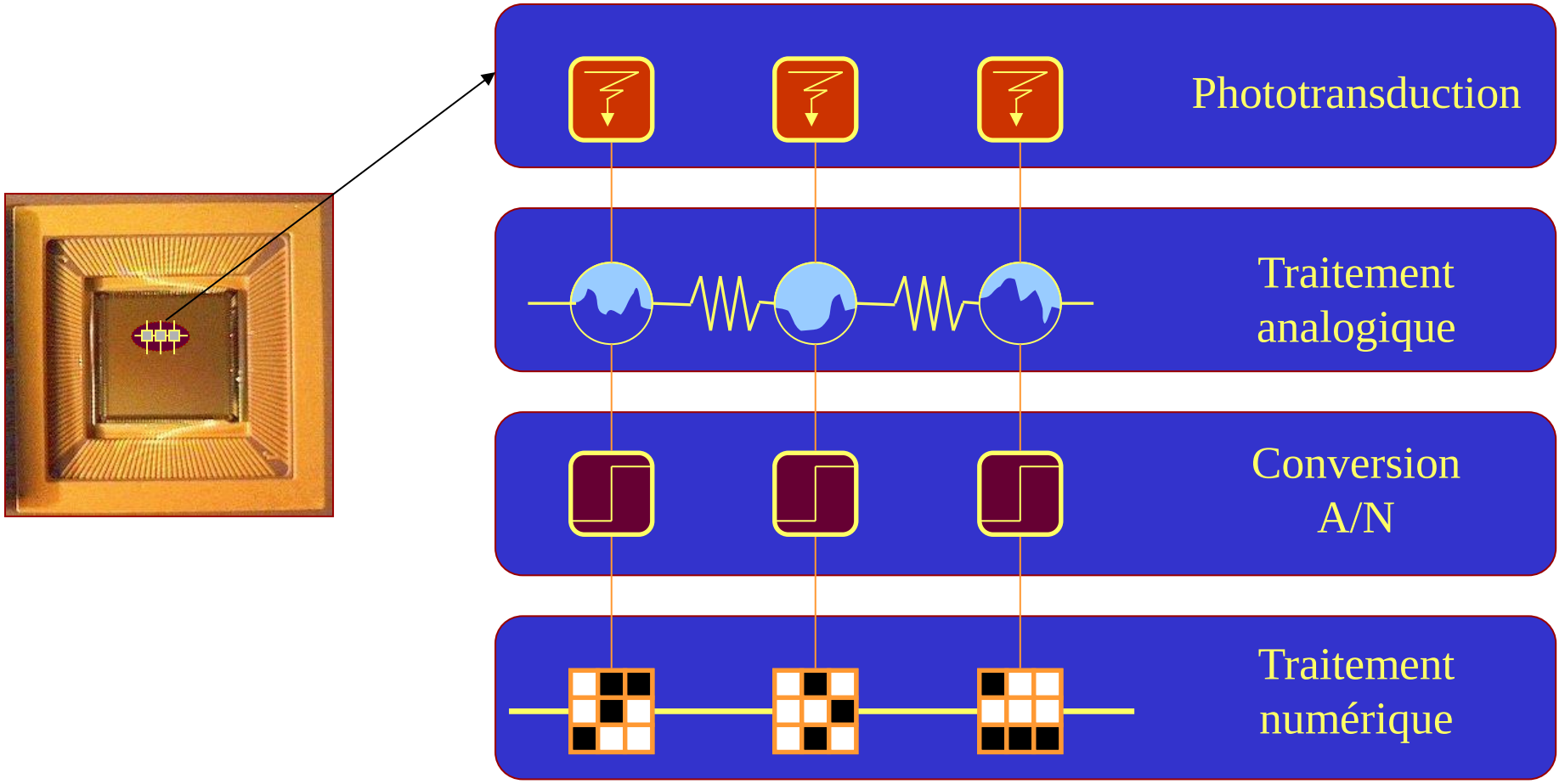
GOULOT D'ETRANGLEMENT DES SYSTEMES DE VISION



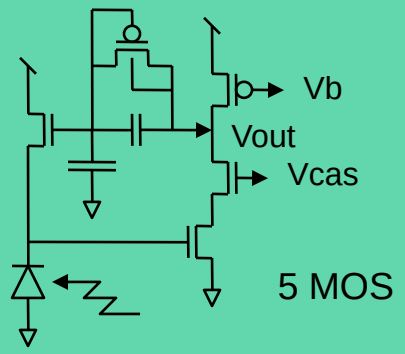
TRAITEMENT D'IMAGE DANS LE PLAN FOCAL



LES RETINES PROGRAMMABLES



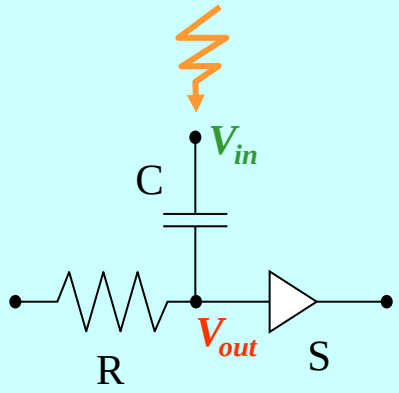
APPROCHES ANALOGIQUES : DETECTION DE MOUVEMENT



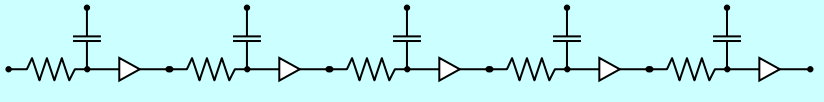
Détection de changement temporel [Delbrück 93]

[Figure Th. Bernard 2007]

APPROCHES ANALOGIQUES : CALCUL DU FLUX OPTIQUE

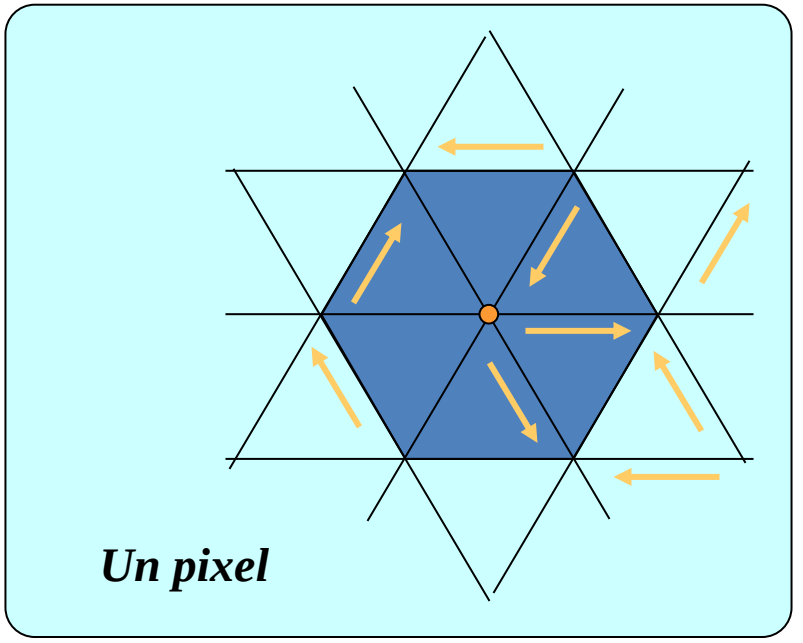


Une cellule de base



Une ligne à retard

- La cellule de base combine un filtre passe-haut du signal d'entrée V_{in} et un filtre passe-bas du signal sortant de la cellule voisine
- La ligne à retard détecte un déplacement qui se produit dans la direction de la ligne.
- Au niveau du pixel, la combinaison des signaux recueilli sur chaque ligne fournit une estimation du vecteur vitesse apparent (flot optique).



Un pixel

[Delbrück 93]

TRAITEMENTS ANALOGIQUES OU NUMERIQUES ?

ANALOGIQUE

☺ Temps continu, asynchrone

☺ Solutions compactes pour opérateurs bas-niveau, linéaire ou non, local ou global

☺ Très basse consommation

☹ Solution figées, ou peu paramétrables

☹ Réutilisabilité (Boîte à outils) difficile

☹ Mise à l'échelle (évolution des technos) problématique

☹ Contrôle et Fiabilité très délicats

NUMERIQUE

☹ Temps discret, synchrone

☹ Coût du séquencement

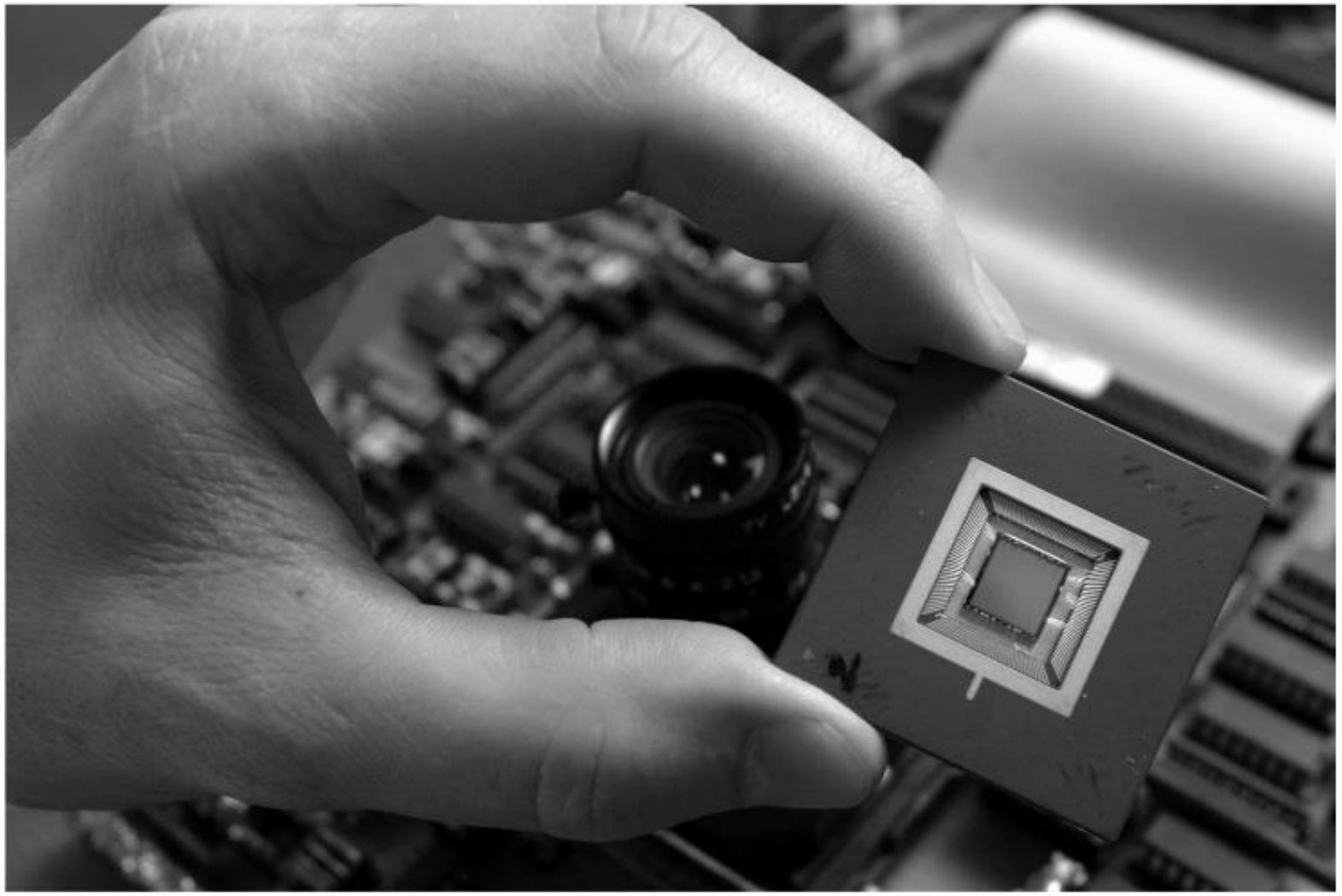
☺ Solutions programmables et évolutives

☺ Construction de boîte à outils aisée

☺ Mise à l'échelle (évolutions des technos) plus simple

☺ Contrôle et Fiabilité mesurables

RETINE PROGRAMMABLE *PVLSAR 34* [T. BERNARD 2004]

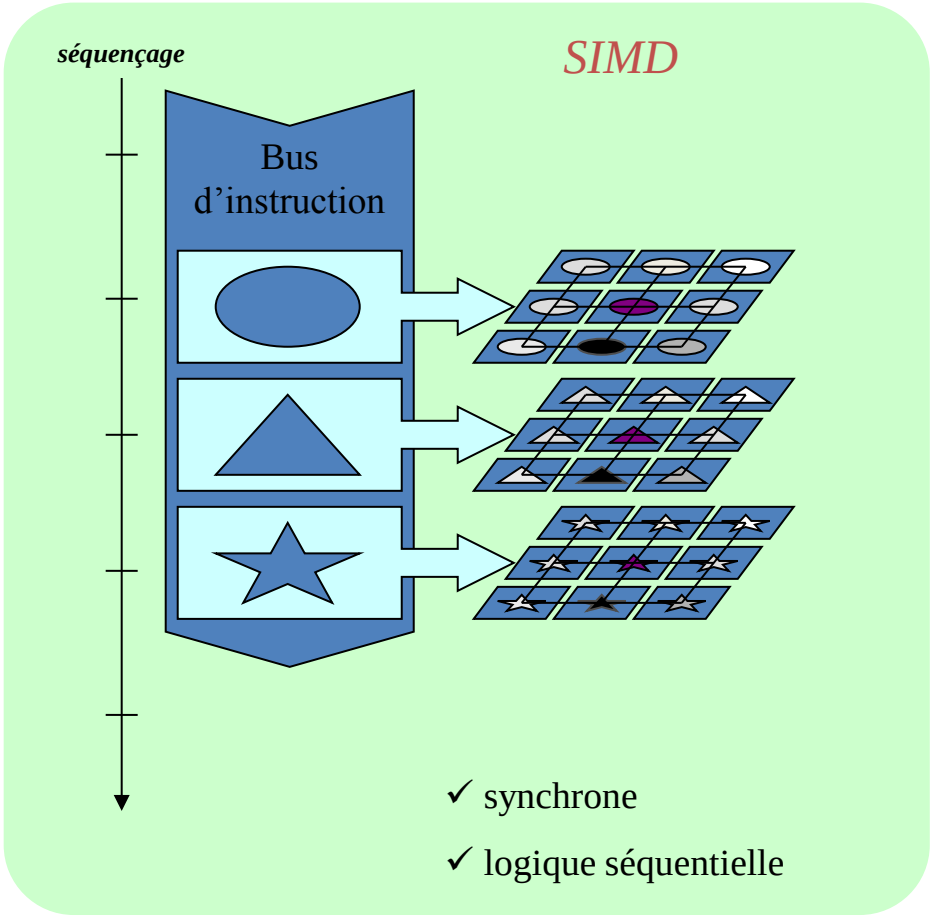


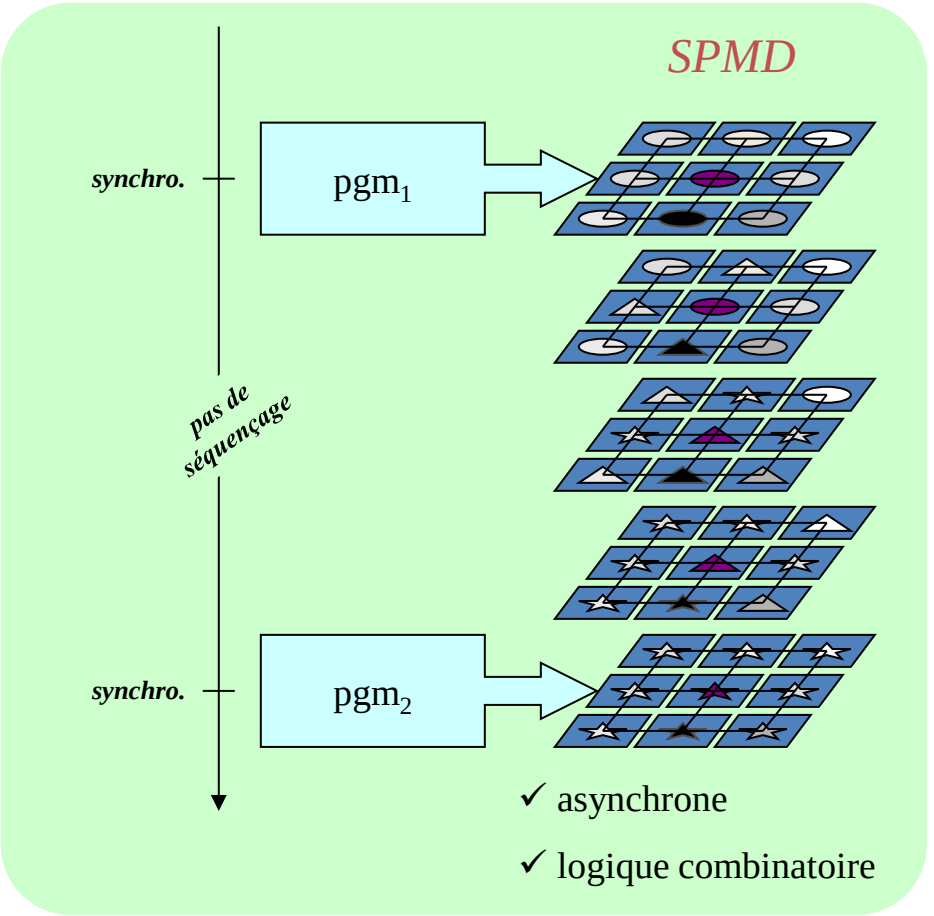
Rétine booléenne
SIMD 200x200

AMS 0,35 µm

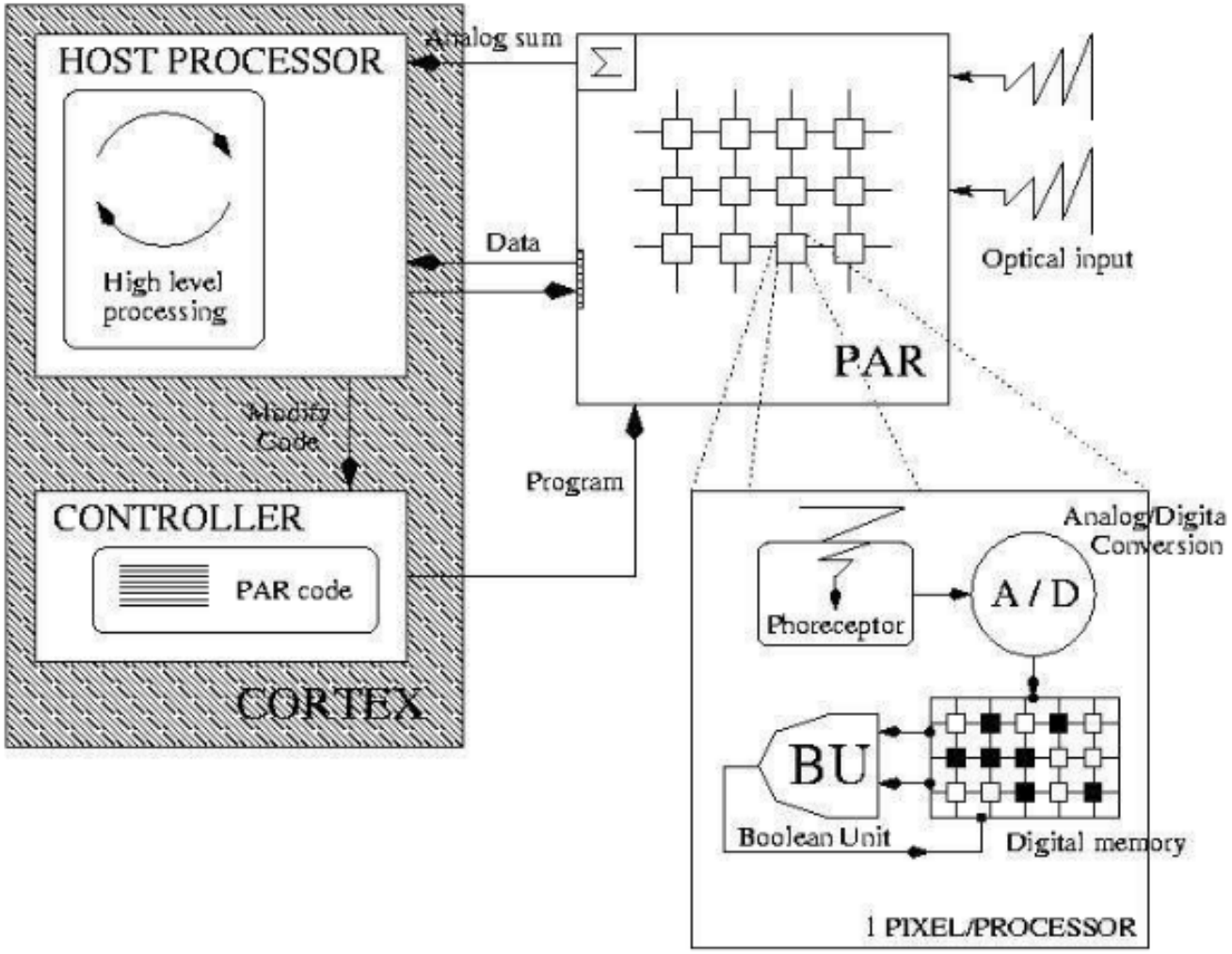
LA RETINE PROGRAMMABLE : MACHINE PARALLELE

Le modèle de fonctionnement de la rétine numérique en tant que *machine parallèle*, est un modèle de type *SIMD* (Single Instruction Multiple Data) ou *SPMD* (Single Program Multiple Data).





SYSTEME DE VISION A BASE DE RETINE



+ Architecture *hétérogène*.

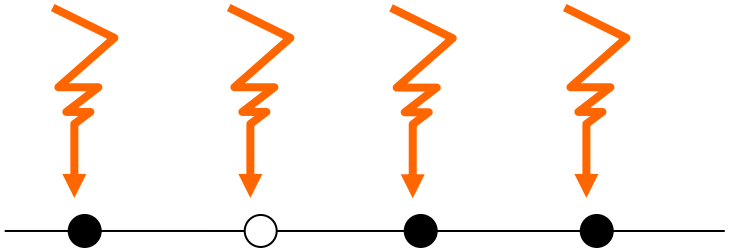
+ Architecture *hybride*.

- *synchrone*
- *asynchrone*
- *numérique*
- *analogique*

+ *Fusion* acquisition/traitement.

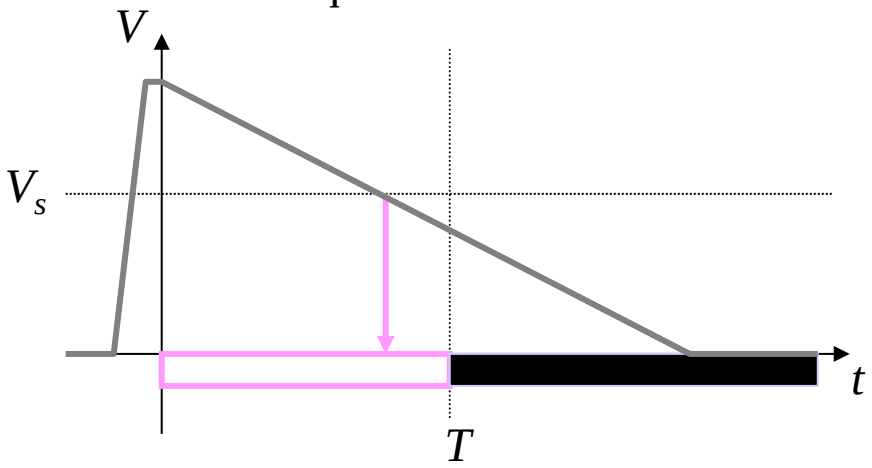
RETOUR EN ARRIERE : RETINE TCL (1993)

1 bit...

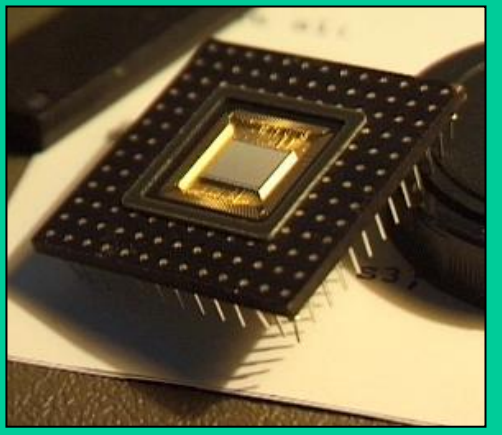


Plan d'acquisition

Acquisition binaire :



Comparaison de la tension aux bornes de la photodiode au temps T par rapport au seuil V_s

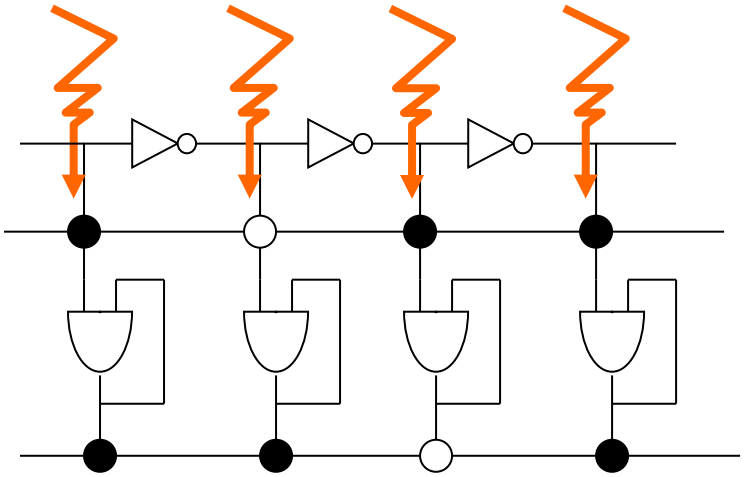


Rétine TCL (Bernard - Zavidovique - Devos 1993)

- ✓ grille de 65x76 pixels
- ✓ CMOS 2 µm
- ✓ 28 transistors/pixel
- ✓ taille du pixel : 100x80 μm^2
- ✓ 1 unité de calcul booléen par pixel
- ✓ mémoire : 3 bits/pixel

RETOUR EN ARRIERE : RETINE TCL (1993)

...2 bits...

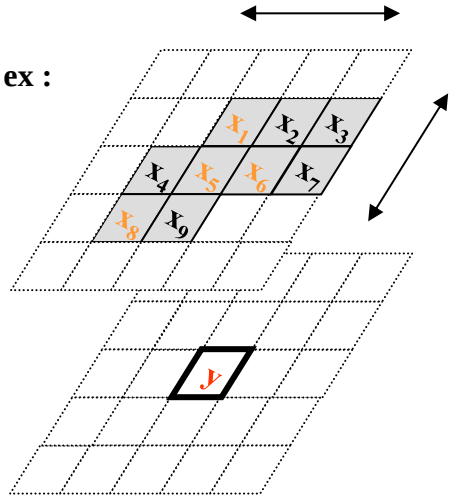


- *Translation (éventuellement complémentée) sur le plan 1*
- *Calcul du ET logique sur le plan 2*

Plan d'acquisition

Calcul de monôme conjonctif :

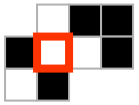
ex :



Plan ET

détection de la présence d'une configuration dans l'image binaire :

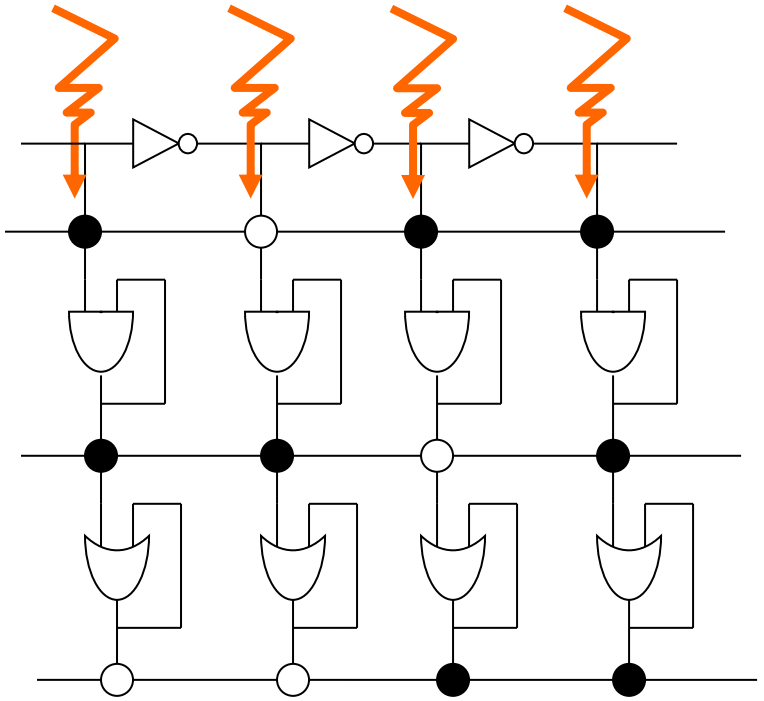
$$y = x_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge \bar{x}_4 \wedge x_5 \; x_6 \wedge \bar{x}_7 \wedge x_8 \wedge \bar{x}_9$$



(Transformée en tout-ou-rien)

RETOUR EN ARRIERE : RETINE TCL (1993)

...3 bits !

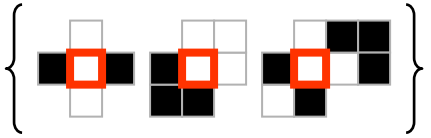


Plan d'acquisition

Plan ET

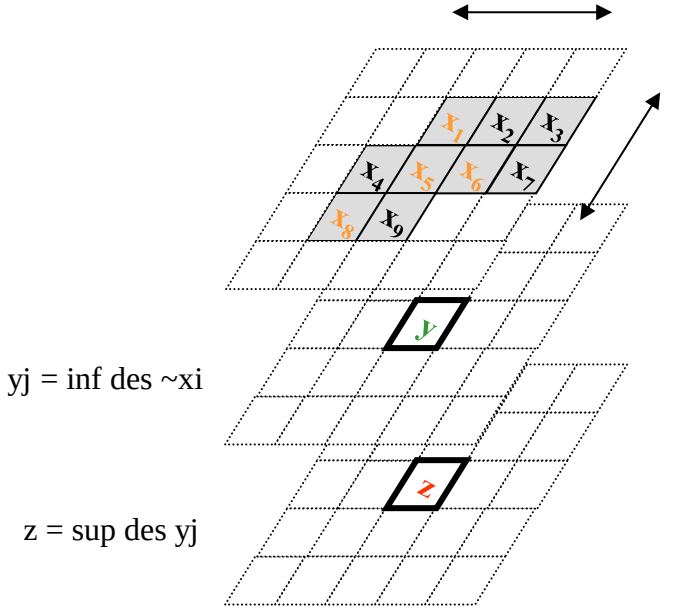
Plan OU

détection de la présence d'une parmi un ensemble de configurations dans l'image binaire :



• *Calcul du OU logique sur le plan 3*

Disjonction de monômes conjonctifs :



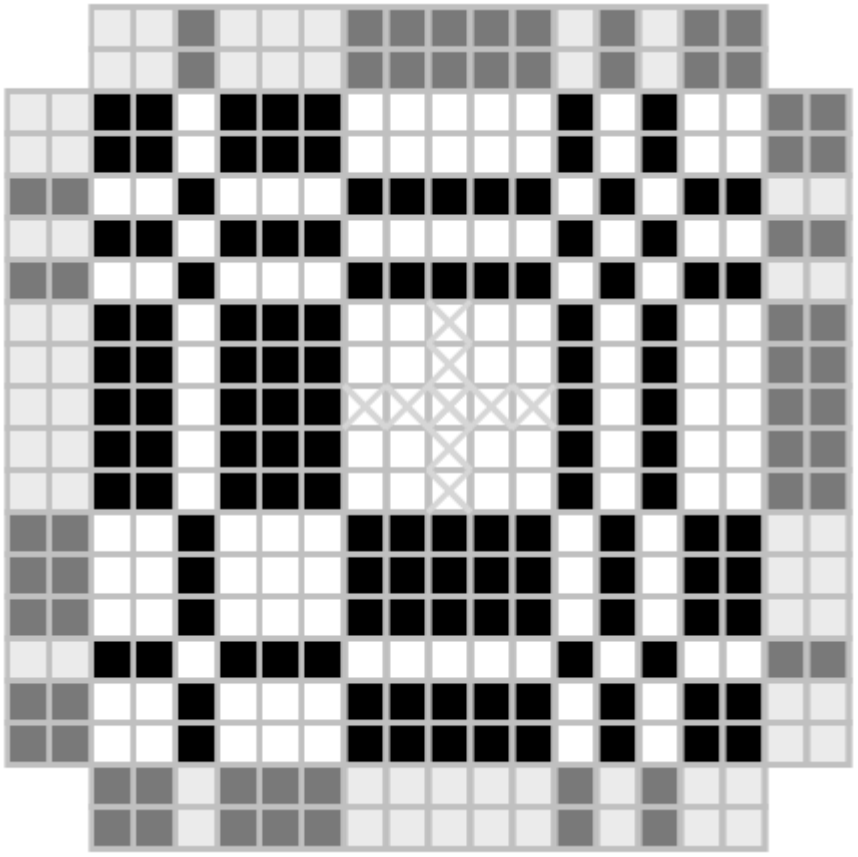
yj = inf des ~xi

z = sup des yj

forme disjonctive

machine booléenne universelle

CODAGE DE POSITION POUR LES RETINES BOOLEENNES



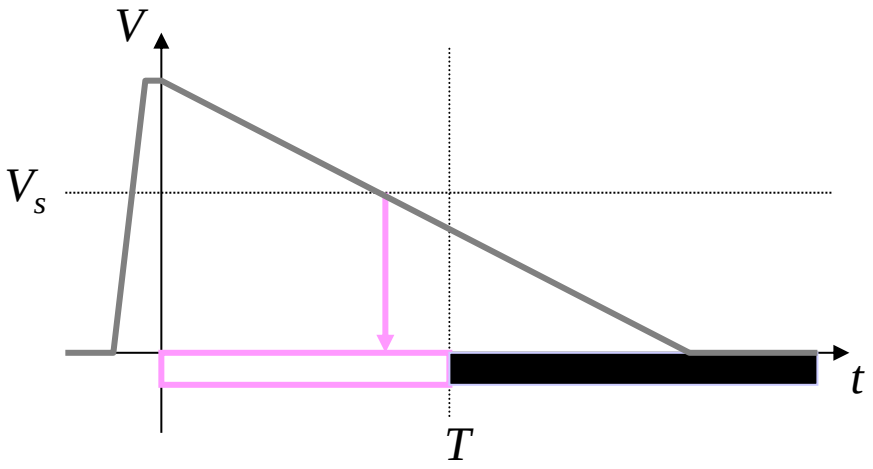
[Bernard 1996]

Grâce aux séquences de De Bruijn 2d, une rétine numérique n'a besoin que d'un bit de mémoire par pixel pour coder localement la position de chaque pixel.

Figure : $B(2,9)$, curseur en croix.

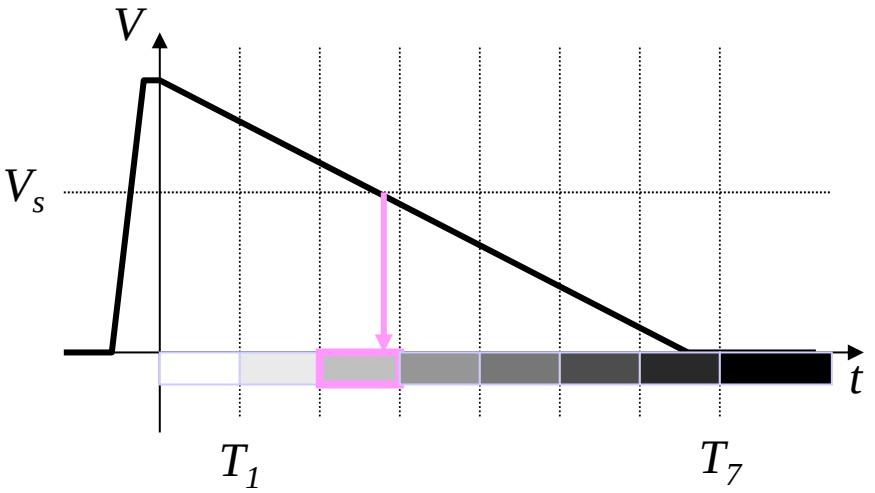
ACQUISITION ET CONVERSION ANALOGIQUE-NUMERIQUE

Acquisition binaire
par seuillage :



Comparaison de la tension aux bornes de la photodiode au temps T par rapport au seuil V_s

Acquisition numérique
par multi-seuillage :

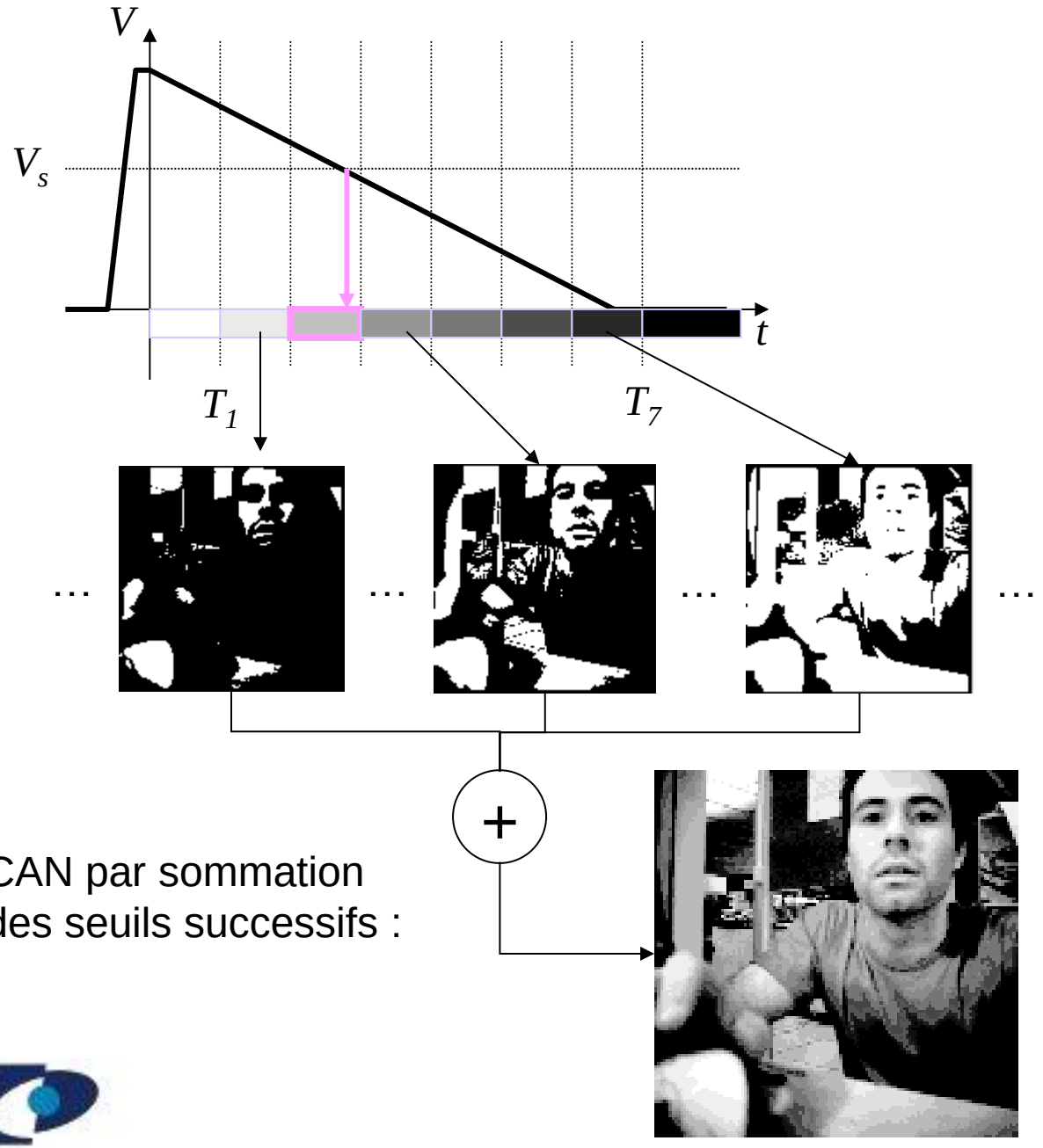


Comparaison de la tension aux bornes de la photodiode aux n temps T_i par rapport au seuil V_s

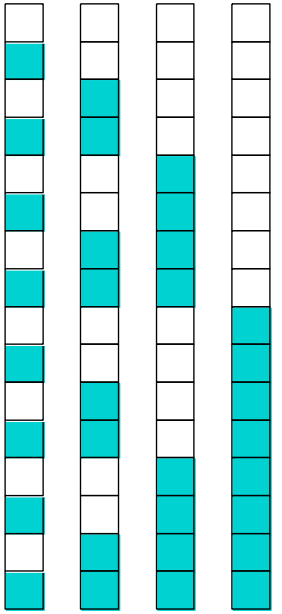
Procédé NSIP
(Near Sensor Image Processing) :

(Eklund - Svensson - Aström 1996)

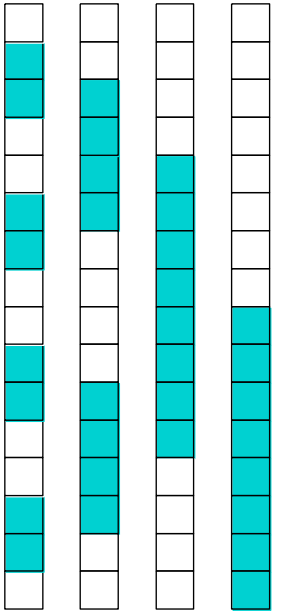
ACQUISITION ET CONVERSION ANALOGIQUE-NUMERIQUE



CAN par sommation des seuils successifs :

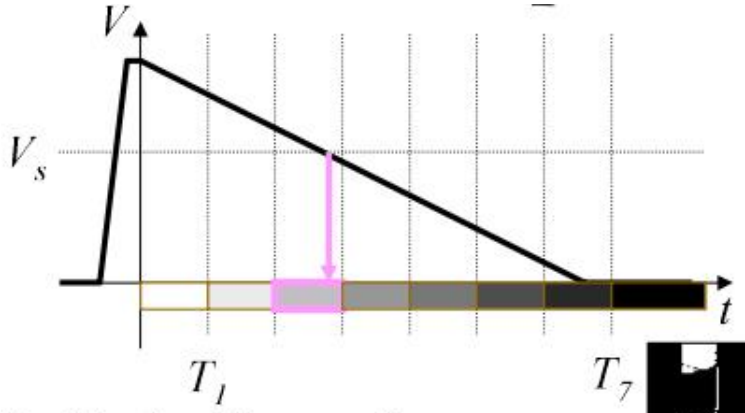


Code naturel : log2(n) opérations par seuil



Code Gray : une seule opération par seuil

FUSION ACQUISITION-TRAITEMENT



Procédé d'acquisition numérique par interrogation multiple du photocapteur au cours du temps

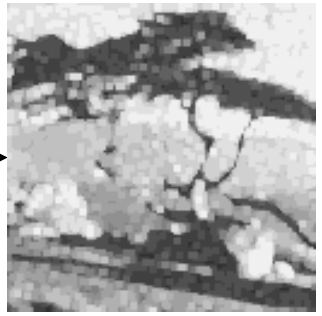
Acquisition active :

- *Adaptation à l'éclairement*
- *Compression logarithmique*
- *Contrôle de gain*

EX. 1 : FILTRES DE RANG NUMERIQUES

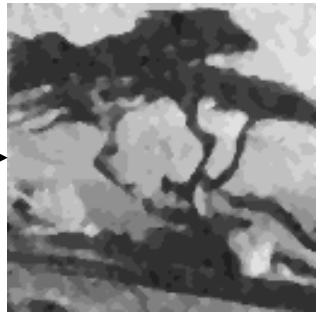
Image bruitée I

T_9^1



dilatation morphologique δ(I)

T_9^5



filtre médian μ(I)

T_9^9

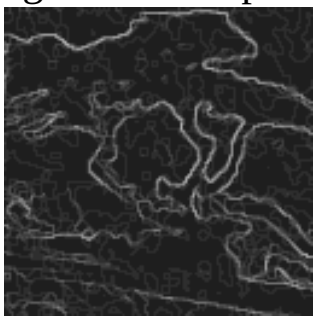
érosion morphologique ε(I)

et par composition/différence :

ouverture morph. δ(ε(I))

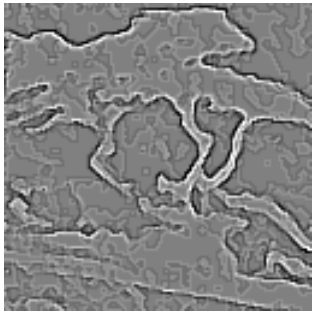
fermeture morph. ε(δ(I))

gradient morph.



μ(I)−ε(μ(I))

laplacien morph.



ε(μ(I))+δ(μ(I))-2μ(I)

EX.2 : CALCUL DES POINTS D'INTERET MORPHOLOGIQUES

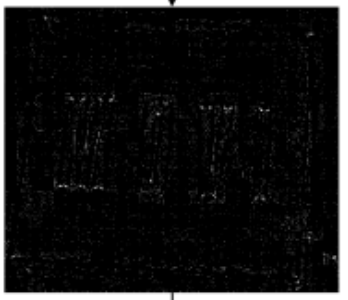
Pour chaque ensemble de niveau :

endo-squelette 8-connexe (courbure positive)

exo-squelette 4-connexe (courbure négative)

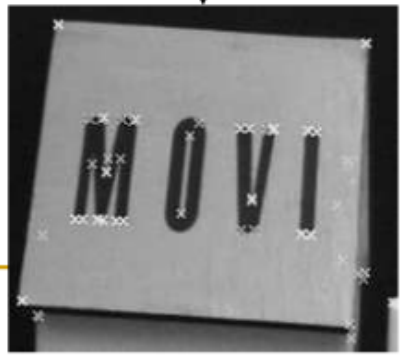
points extrêmes

somme



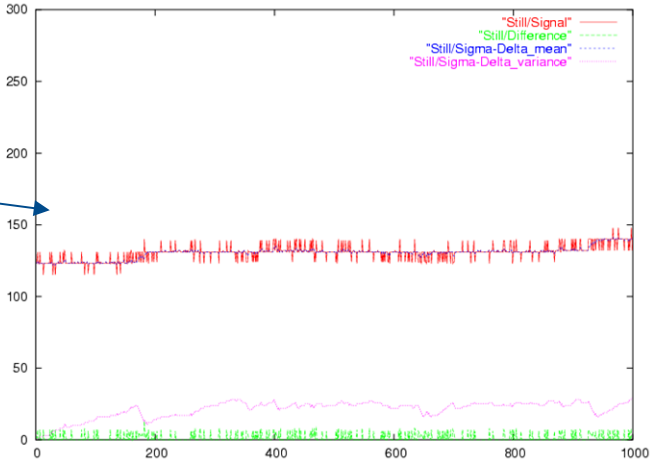
Fonction d'intérêt : produit de la courbure par le module du gradient.

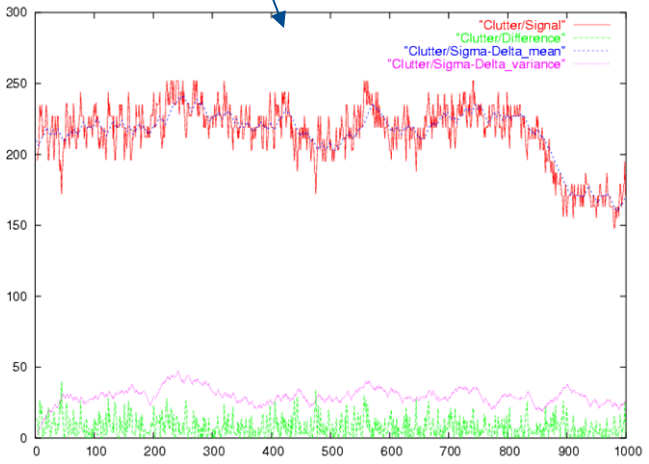
Maxima locaux

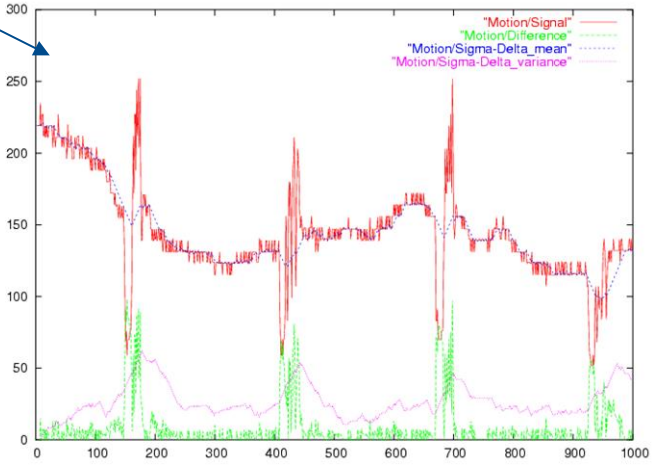


points d'intérêt

DETECTION D'OBJETS MOBILES SUR *PVLSAR 34…*







…voir la suite dans le cours « Détection de mouvement » !!!

COURS N°2 : CONCLUSIONS

Les techniques de co-conception visent à optimiser globalement un système de vision par une démarche opportuniste qui exploite les différents éléments du système et cherche à les combiner de façon plus intime : optique, mécanique, photo-capture, numérisation, traitement…

Ce cours s'est concentré sur la vision 3d et l'analyse du mouvement (rétines). On trouvera encore plus d'exemples de co-conception de caméras dans le domaine de la "photographie computationnelle" (voir Conférence *DxO Labs*), pour l'amélioration et l'"augmentation" des images numériques.

On retiendra l'équilibre entre, d'une part la complexité matérielle et le caractère intrusif (éclairage) du système, et d'autre part la complexité logicielle.

Le poids du logiciel reste néanmoins important sur la plupart des systèmes présentés.

On notera, pour les systèmes passifs, la difficulté, voire l'impossibilité de traiter les zones dépourvues de structures (homogènes).

L'origine des principes présentés est souvent ancienne, mais la matûration des technologies est très récente, avec plusieurs produits sur étagères disponibles aujourd'hui.

REFERENCES (1ère Partie)

[Chiabrando 2009] F. Chiabrando, R. Chiabrando, D. Piatti, F. Rinaudo, *Sensors for 3D Imaging: Metric Evaluation and Calibration of a CCD/CMOS Time-of-Flight Camera*, Sensors, vol. *9*, 10080-10096, 2009.

[Geng 2011] Jason Geng, *Structured-light 3D surface imaging: a tutorial*, Advances in Optics and Photonics, vol. *3*, 128-160, 2011.

[Posdamer 1982] J. L. Posdamer and M. D. Altschuler, *Surface measurement by space-encoded projected beam systems*, Comput. Graph. Image Processing 18, (1), 1–17 1982.

[Narasimhan 2006] S. Narasimhan, *Computer Vision: Spring 2006, lecture n.17*, Carnegie Mellon University.

[Zhang 2002] L. Zhang, B. Curless, S. M. Seitz, *Rapid shape acquisition using color structured light and multi-pass dynamic programming*, IEEE Int. Symp. on 3D Data Processing Visualization and Transmission, pp. 24–36, 2002.

REFERENCES (2ème Partie)

[Adelson 1992] E. H. Adelson, J. Y. A. Wang, *Single Lens Stereo with a Plenoptic Camera*, IEEE Trans. Pattern Analysis and Machine Intelligence 14(2): 99-106, 1992.

[Ng 2005] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan. *Light Field Photography with a Hand-Held Plenoptic Camera,* Stanford University Computer Science Tech Report CSTR 2005-02, April 2005.

[Pentland 1987] Alex P. Pentland, *A new sense for depth of field*, IEEE Trans. Pattern Analysis and Machine Intelligence 9(4): 523-531, 1987.

[Maître 2003] Henri Maître (ss la direction de), *Le Traitement des Images,* Chapitre 5 : Restauration, Hermès – Lavoisier, Série I2C, 2003.

[Levin 2007] A. Levin, R. Fergus, F. Durand, W.T. Freeman, *Image and depth from a conventional camera with a coded aperture,* ACM Transactions on Graphics 26 (3): 70-78, 2007.

REFERENCES (3ème Partie)

[Delbruck 1993] Toby Delbrück, *Silicon retina with Correlation-Based, Velocity-Tuned Pixels*, IEEE Transactions on Neural Networks, Vol. 4, No. 3, pp. 529–541, 1993.

[Bernard 1993] T.M. Bernard, B. Zavidovique, and F.J. Devos, *A programmable artificial retina*, IEEE Journal of Solid-State Circuits, 28(7), Jul 1993, p. 789-798.

[Bernard 1996] T.M. Bernard and J.C. Meier, *Cursor-Injective Two-Valued Lattices for a Local Encoding of Pixel Position*, Proc. SPIE, Vol. 2950, Advanced Focal Plane Arrays and Electronic Cameras, 230-241,1996.

[Astrom 1996] A. Aström, R. Forchheimer, and J.E. Eklund, *Global feature extraction operations for near-sensor image processing*, IEEE Transactions on Image Processing, 5(1), 102-110, 1996.

[Lacassagne 2009] L. Lacassagne, A. Manzanera, J. Denoulet, and A. Mérigot, *High performance motion detection: some trends toward new embedded architectures for vision systems.* Journal of Real Time Image Processing, 4(2), 2009, pp. 127--146.

QUELQUES SUGGESTIONS POUR LES EXPOSES ORAUX…

- Systèmes de vision inspirés de la **vision des abeilles ou des mouches**, ex : équipe Biorobotique de l'institut des sciences du mouvement (Jules Marey), à Marseille…
- **Temps avant collision** : implantations dédiées, ou études biologiques…
- **Mouvement et Gestalt** : d'autres exemples de groupement perceptuels ou simplification, liens avec d'autres aspects du Gestalt…
- **Accomodation /Autofocus** : mécanismes biologiques, opto-mécaniques, algorithmiques…
- **Random dot (auto)stereograms** : création, appariement, lien avec les textures…
- **Perspective et Gradients de Texture** : techniques de géométrie descriptive pour le dessin, et/ou reconstruction 3d à partir de la perspective.
- Systèmes de vision active utilisant une exploration inspirée par les **mouvements oculaires** humain.
- Systèmes de **super-résolution** fondés sur les micro-mouvements (micro-saccades)…
- **Masquage saccadique** : utilisation dans un système de vision active…
- Mesure du **déphasage par obturateur** pour une caméra temps de vol, ex : Kinect v2…
- **Utilisation des ombres ou de l'éclairage « naturel »** comme lumière structurée pour la reconstruction 3d…
- Lumière structurée : propriété des **mires 2d pseudo-aléatoires**…
- **Images et écrans lenticulaires** pour l'affichage 3d ou l'acquisition plénoptique…
- Utilisation de l'**aberration chromatique** pour résoudre l'ambiguïté de la profondeur par le défocus (travaux Pauline Trouvé et al, 2013)
- **Translation du plan focal** pendant l'acquisition pour augmenter la profondeur de champ (travaux Hajime Nagahara, 2008)