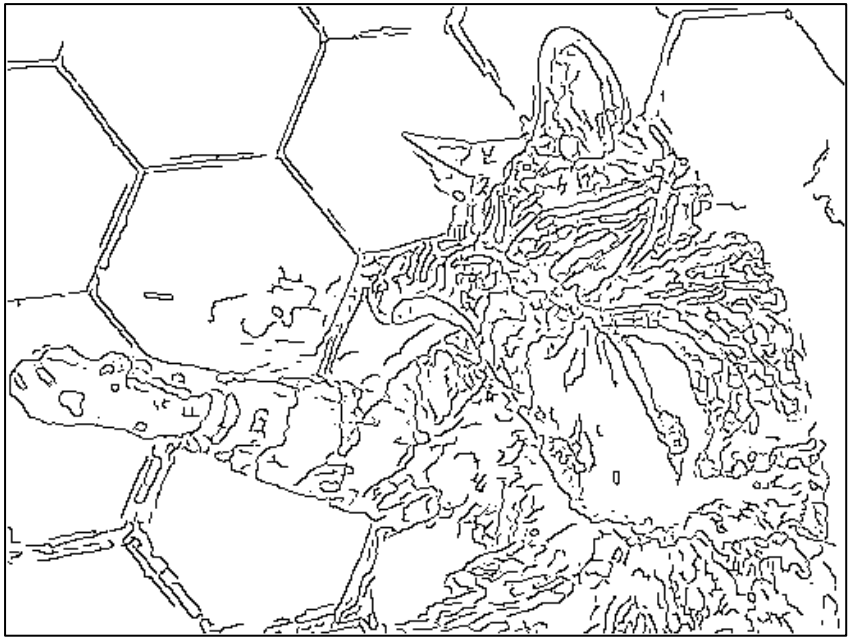
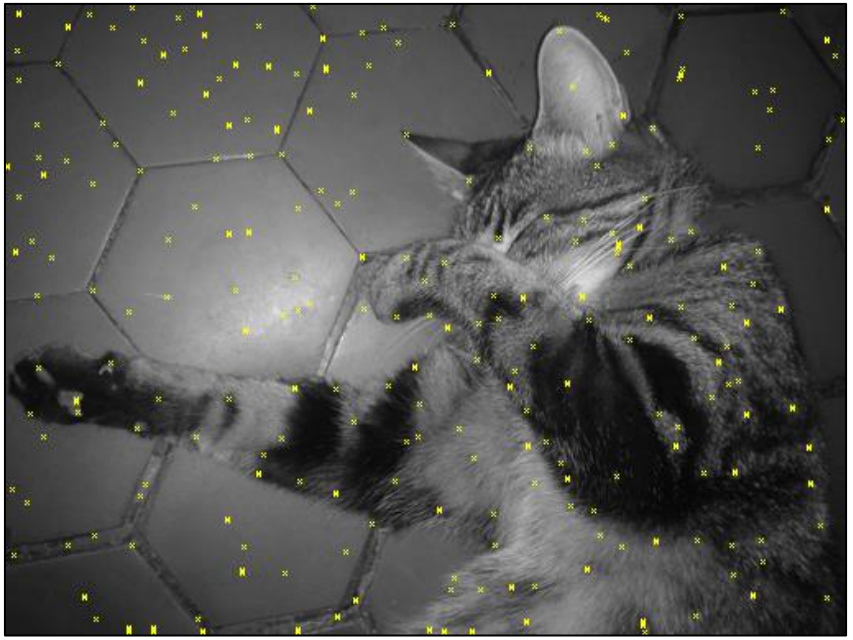
# Indexation d’images

## *CARACTERISTIQUES MULTI-ECHELLES*





**Antoine Manzanera - ENSTA-ParisTech / U2IS**

**Cours Master AIC – Université Paris Saclay**

ENSTA ParisTech

# CARACTERISTIQUES VISUELLES MULTI-ECHELLES

Les caractéristiques visuelles ont pour objet de **représenter** les objets afin de faciliter leur **appariement** dans les images (séquences, paires, bases, modèles,…)

L'extraction de caractéristiques dans les images consiste à :

1) **Réduire le support** de la représentation dans les images à un sous-ensemble **significatif** et **parcimonieux**.
2) Calculer une **fonction décrivant** ce sous-ensemble de façon **discriminante**, **robuste** et **efficace**.

La caractérisation **locale** est en général liée à la **géométrie** locale (différentielle).

La caractérisation **globale** est en général liée à une représentation **statistique**.

Le calcul multi-échelles permet de :

1) Fournir une **base formelle** correcte au calcul différentiel.
2) Etablir un **continuum** entre le local (géométrique) et le global (statistique).

# CARACTERISTIQUES MULTI-ECHELLES

Plan du cours :

- Introduction : qu'est-ce qu'une bonne primitive visuelle ?
- Bases de géométrie différentielle pour les images
- Du local au régional : les dérivées multi-échelles
- Détection de contours multi-échelles
- Points d'intérêt 1 : détecteur de Harris
- Points d'intérêt 2 : points SIFT
- Descripteurs locaux 1 : invariants de Hilbert
- Descripteurs locaux 2 : Histogrammes d'orientation
- Du local au global : Sacs de mots visuels
- Descripteurs globaux : Invariants de Fourier-Mellin

# QU'EST-CE QU'UNE BONNE PRIMITIVE VISUELLE ?

**Objectif** : mettre en correspondance des points / ensembles / images avec d'autres points / ensembles / classes / catégories visuelles.

Les propriétés souhaitées d'une bonne primitive sont :

- **Robustesse** : L'information visuelle doit être fidèlement représentées indépendamment des modifications qu'elle peut subir d'une instance à l'autre : distorsions géométriques, changements d'illumination, occultations, variations intra-classes…
- **Discrimination** : L'objet représenté doit se distinguer facilement des autres objets, en particulier de ceux qui l'environnent.
- **Efficacité** : le calcul de la primitive doit être rapide, et les descripteurs économes en taille mémoire…

Une primitive visuelle peut caractériser l'information image à plusieurs niveaux :

- **Local** : une primitive par point / région / courbe…
- **Global** : une primitive qui concerne l'ensemble de l'information…

# BASES DE GEOMETRIE DIFFERENTIELLE POUR LES IMAGES

La « géométrie locale » dans une image se décrit naturellement par des concepts de la géométrie différentielle : direction, courbure,…

Dans le modèle différentiel, l'image est assimilée à une fonction $I: \mathbb{R}^2 \longrightarrow \mathbb{R}$ continue et différentiable.

Le comportement local de l'image autour de chaque point peut être prédit par ses dérivées locales (Formule de Taylor) :

$$I(x_0+\varepsilon, y_0+\eta) = \sum_{k=0}^{r}\sum_{i=0}^{k} C_k^i \varepsilon^{k-i}\eta^i \frac{\partial^k I}{\partial x^{k-i}\partial y^i}(x_0,y_0) + o\left((\varepsilon^2+\eta^2)^{r/2}\right)$$

Dans les images discrètes, la notion de dérivabilité sera remplacée par la notion de régularité locale.

Cette régularité pouvant être imposée explicitement par filtrage (convolution), l'estimation de la dérivée correspondra à une convolution, et sera toujours relative à une échelle (espaces d'échelles).

# ORDRE 1 : GRADIENT ET ISOPHOTE

A l'ordre 1, la grandeur de base est le vecteur gradient :

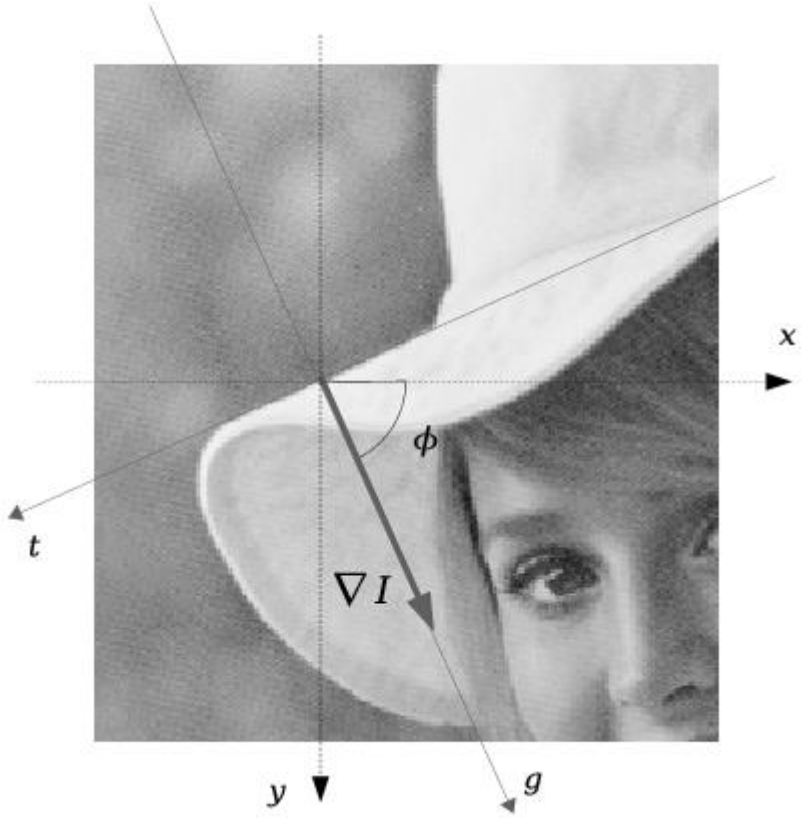
$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right)^T$$

- Son argument, $\arg \nabla I$, correspond à la direction de plus grande pente.
- Sa norme, $\|\nabla I\|$, mesure le contraste local.
- Il permet de calculer la dérivée partielle dans toute direction du plan. Soit v un vecteur unitaire :

$$\frac{\partial I}{\partial v} = \nabla I . v^T$$

- Dans le repère local (*g*,*t*) avec $g = \frac{\nabla I}{\|\nabla I\|}$ et $t = g^{\perp}$ :

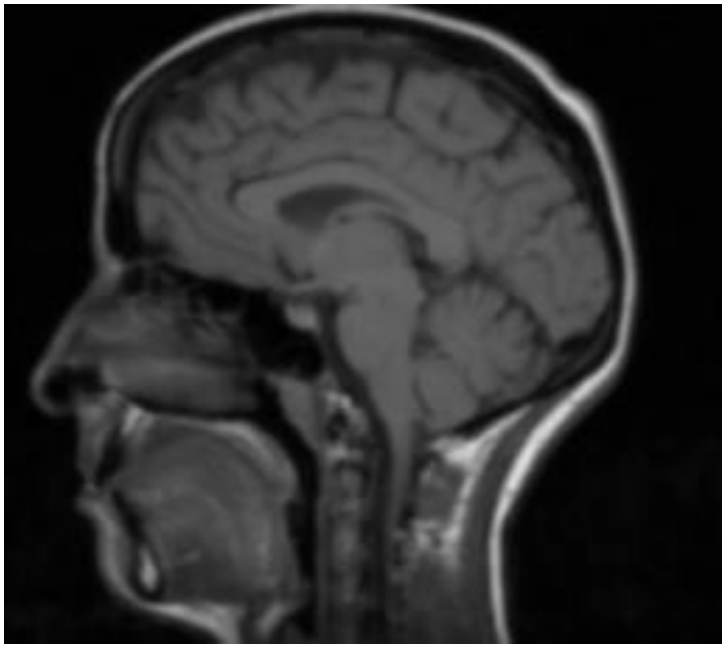
$\frac{\nabla I}{\nabla g} = \left|\|\nabla I\|\right|$ (direction principale) ; $\frac{\nabla I}{\nabla t} = 0$ (isophote)

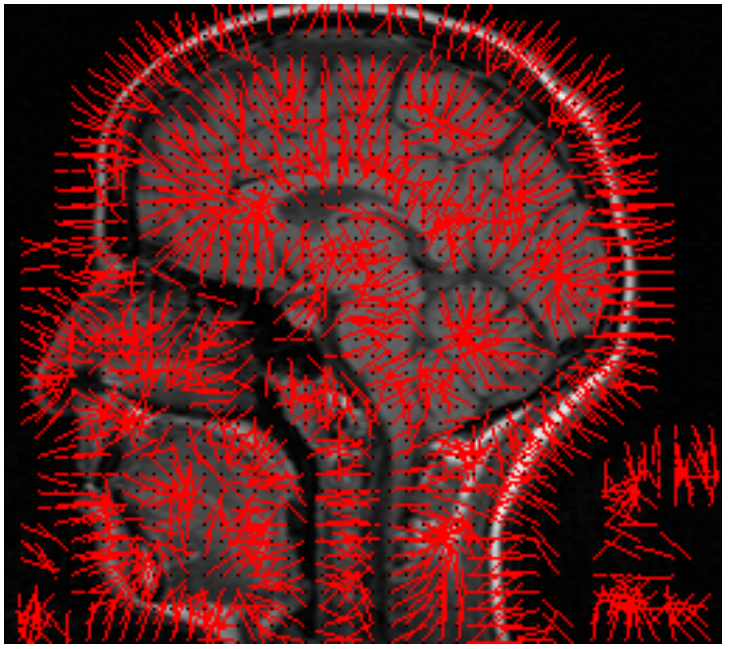


# GRANDEURS DIFFERENTIELLES D'ORDRE 1

*original*

$I$



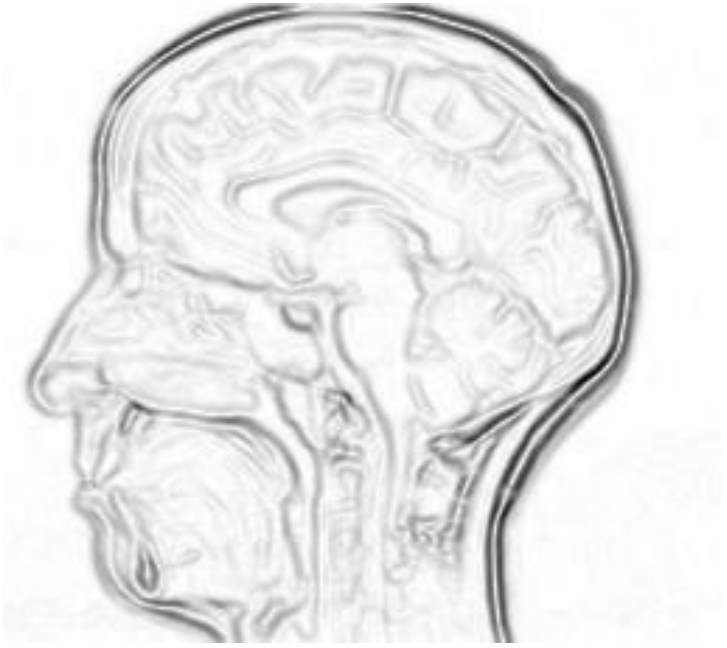


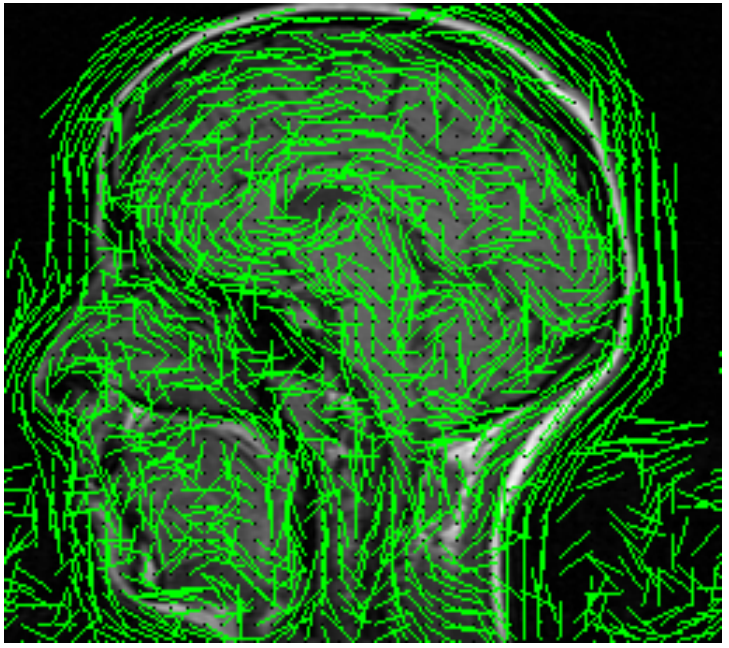
*direction du gradient*

$\arg \nabla I$

*norme du gradient*

$\|\nabla I\|$





*direction de l'isophote*

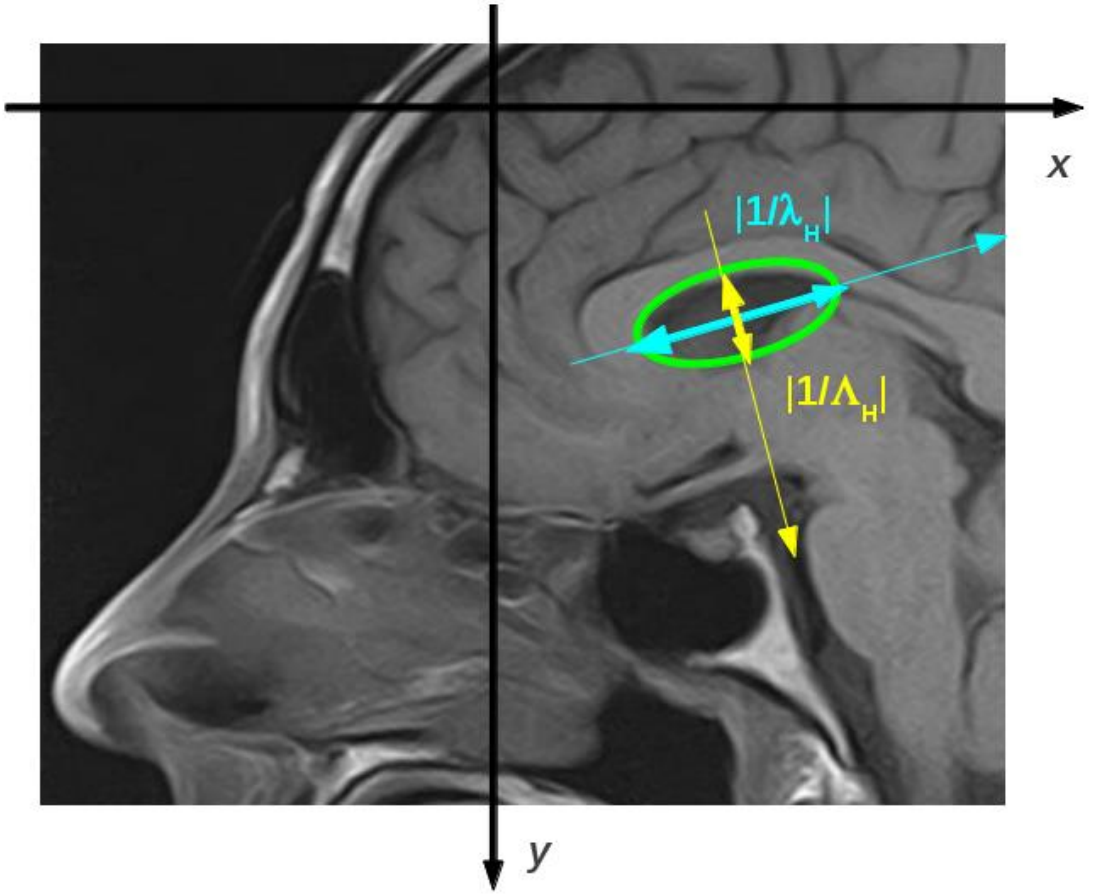
$\arg \nabla I^{\perp}$

# ORDRE 2 : HESSIENNE ET COURBURE

A l'ordre 2, la grandeur de base est la matrice hessienne :

- Ses vecteurs propres (resp. ses valeurs propres $\Lambda_H$ et $\lambda_H$) correspondent aux directions (resp. intensités) de courbure principale.
- Sa norme de Frobénius, $\|\mathrm{H_I}\|_F$ , mesure l'intensité de la courbure globale.

$$H_I = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{pmatrix}$$



# ORDRE 2 : HESSIENNE ET COURBURE

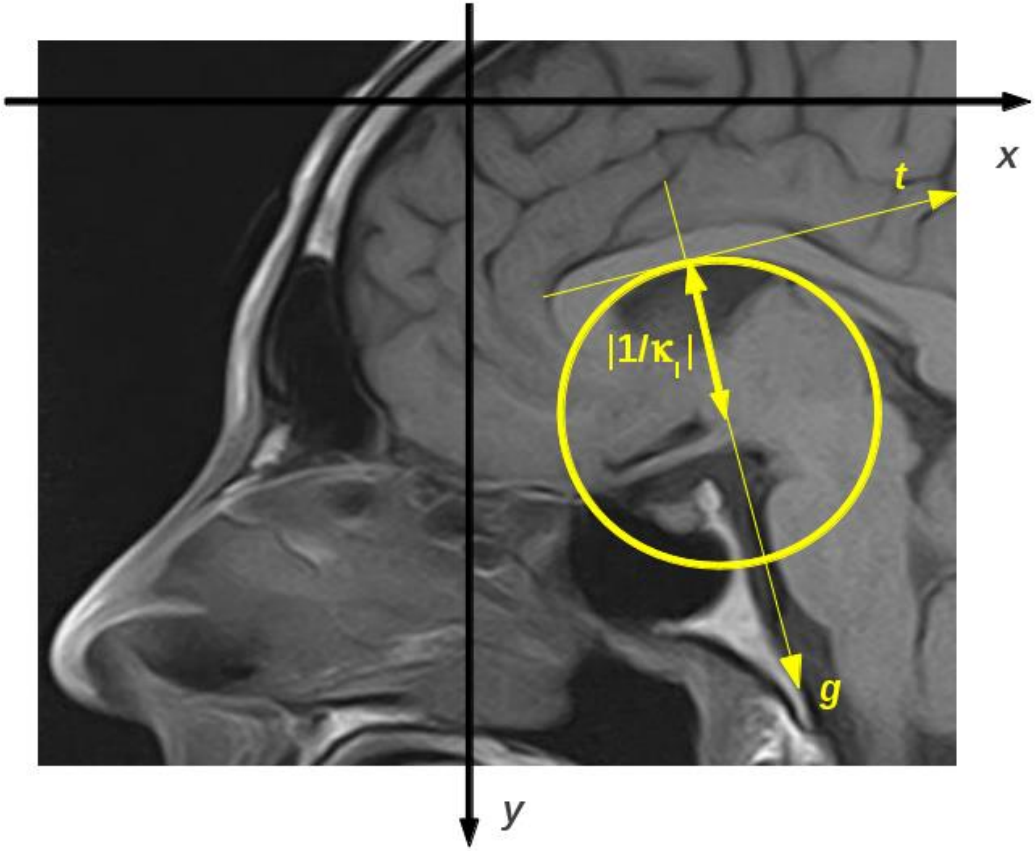
- Soit *u* et *v* deux vecteurs unitaires. La dérivée seconde selon u et v se calcule par :

$$\frac{\partial^2 I}{\partial u \partial v} = u^T H_I v$$

- En particulier la courbure de l'isophote est égal à l'inverse du rayon du cercle osculateur au contour :

$$\kappa_I = -\frac{I_{tt}}{I_g} = -\frac{I_{xx}I_y^2 - 2I_xI_yI_{xy} + I_{yy}I_x^2}{\|\nabla I\|^3}$$

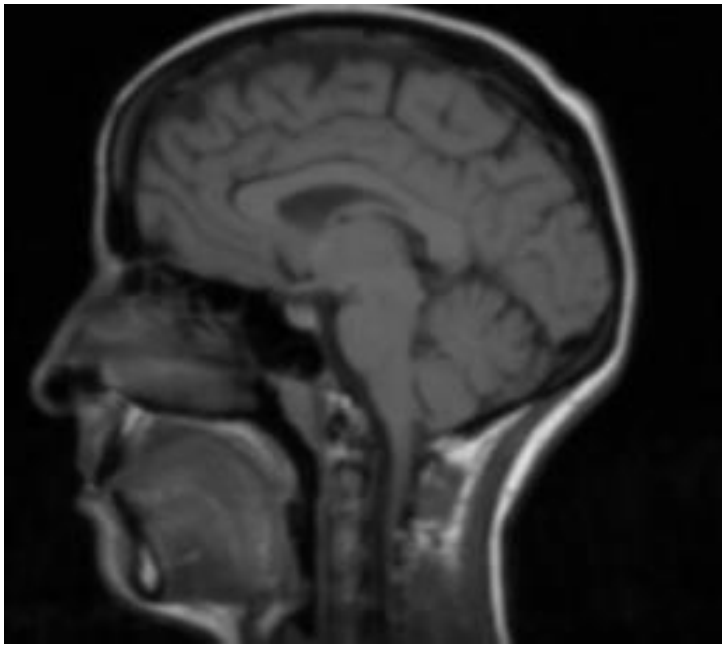
(Notations :
$I_u = \frac{\partial I}{\partial u}; I_{uv} = \frac{\partial^2 I}{\partial u \partial v}$, etc.)

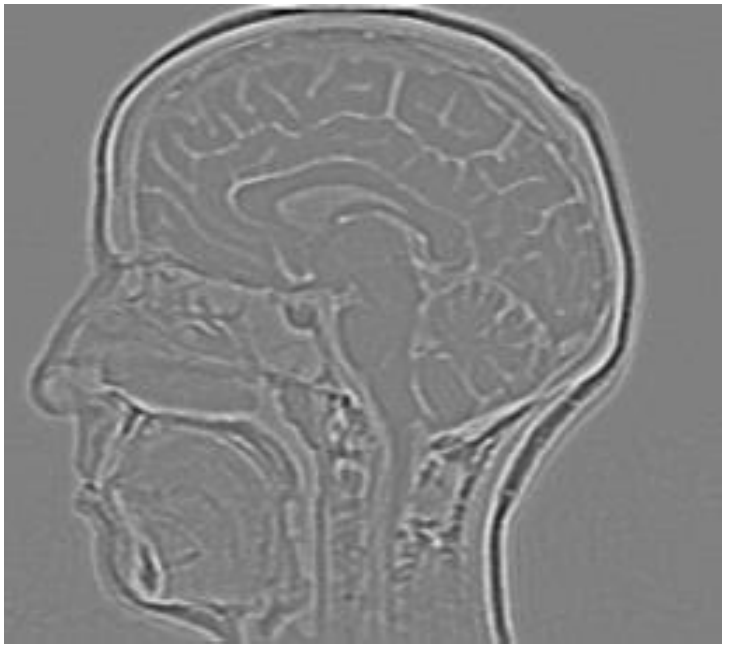


# GRANDEURS DIFFERENTIELLES D'ORDRE 2

*original*

$I$



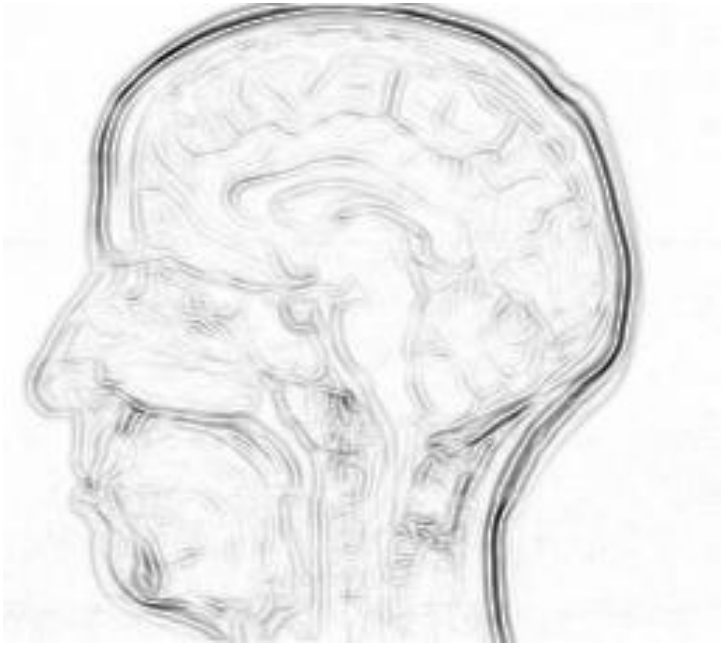


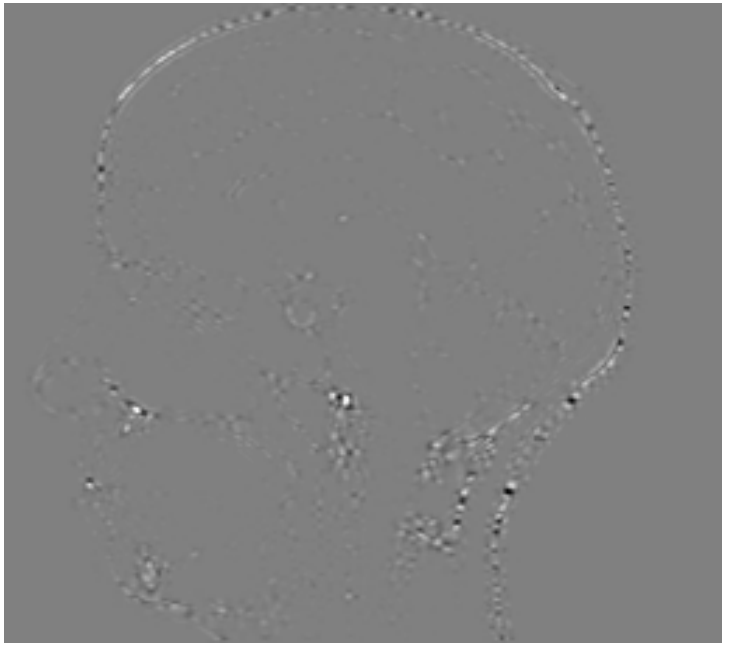
*trace de la hessienne, ou courbure totale = laplacien*

$\Delta I$

*norme de la hessienne*

$\|\mathrm{H_I}\|_F$





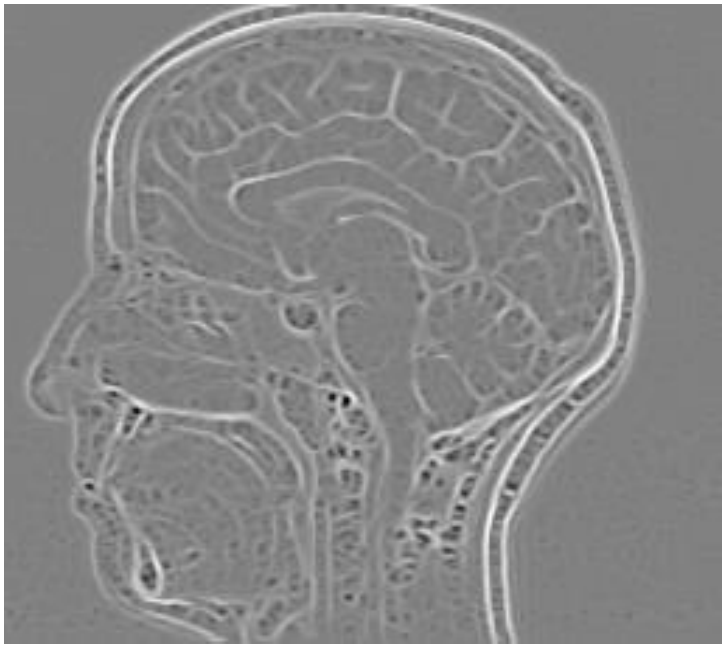
*déterminant de la hessienne*

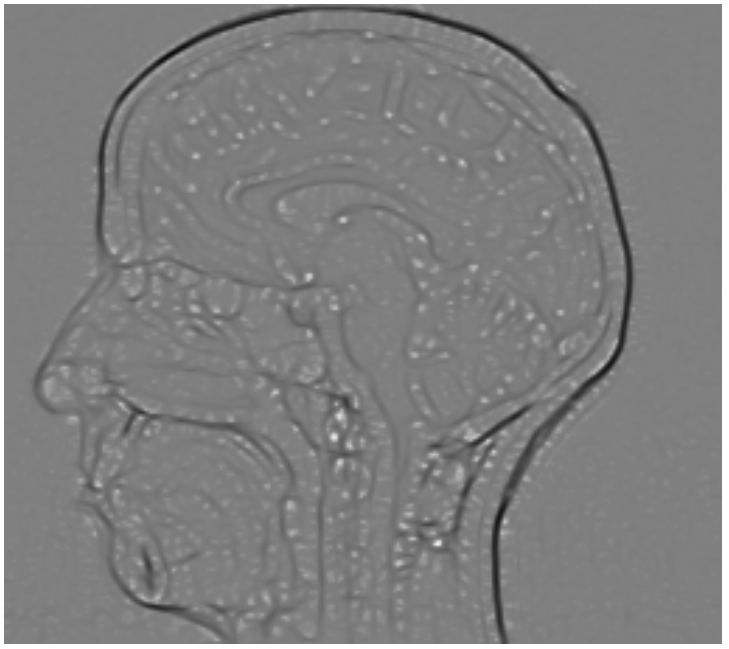
$\det\|\mathrm{H_I}\|_F$

# GRANDEURS DIFFERENTIELLES D'ORDRE 2

*« grande » valeur propre*

$\Lambda_I$

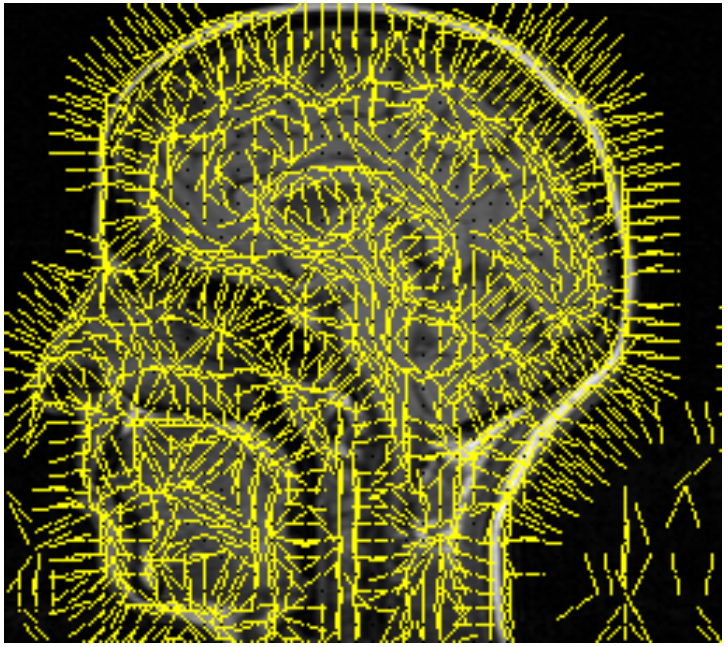


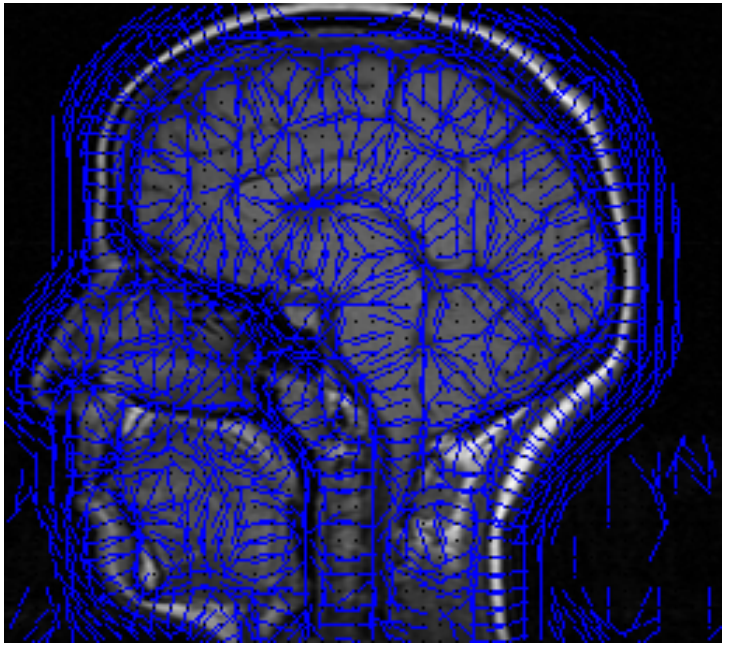


*« petite » valeur propre*

$\lambda_I$

*direction du « grand » vecteur propre*





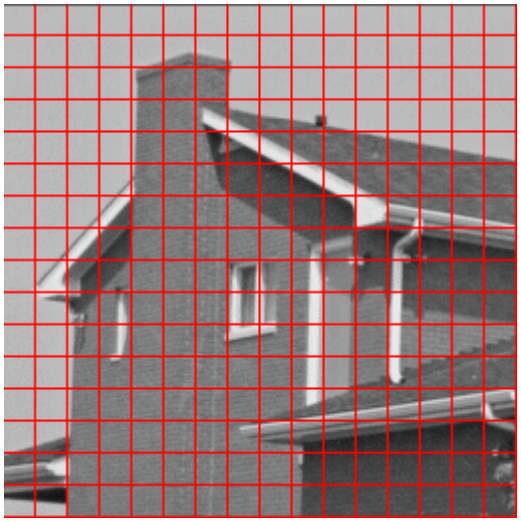
*direction du « petit » vecteur propre*

# REPRESENTATION PAR LES DERIVEES LOCALES

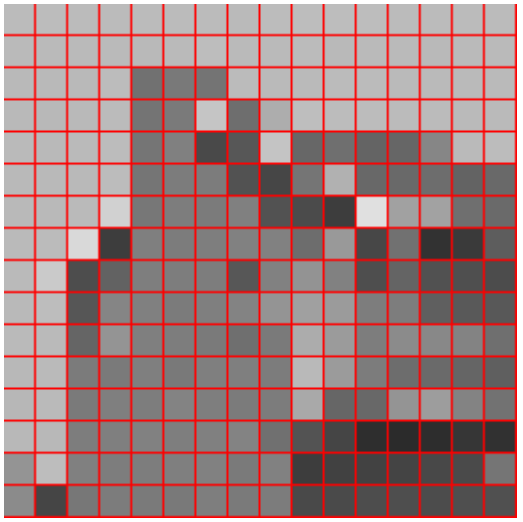
Expression de la formule de Taylor à l'ordre 2, à partir du vecteur gradient et de la matrice hessienne :

$$I(x_0 + \varepsilon, y_0 + \eta) = I(x_0, y_0) + (\varepsilon, \eta)^T . \nabla I + (\varepsilon, \eta)^T . H_I . (\varepsilon, \eta) + o(\varepsilon^2 + \eta^2)$$

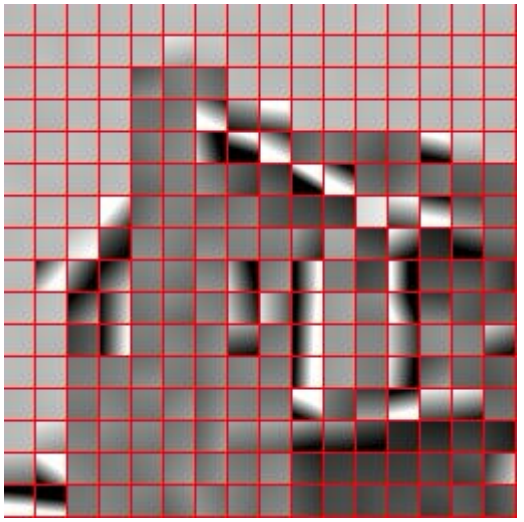
Reconstruction d'imagettes à partir des dérivées partielles calculées au centre de l'imagette, à l'ordre 0, 1 et 2 :



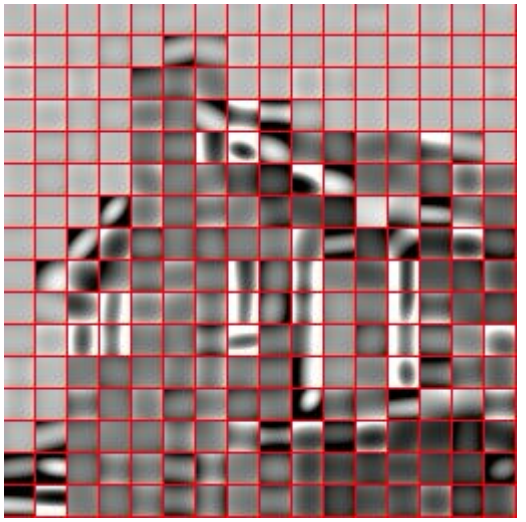
*imagettes*



*ordre 0*



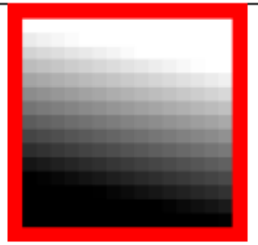
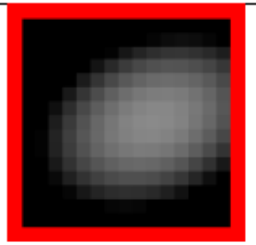
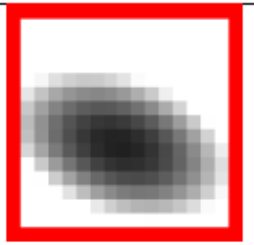
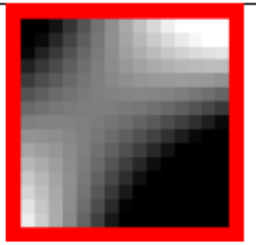
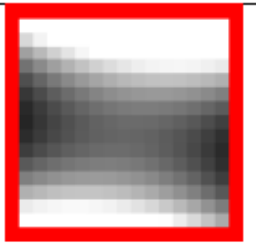
*ordre 1*



*ordre 2*

# CATEGORISATION DIFFERENTIELLE LOCALE

La valeur des dérivées jusqu'à l'ordre 2 permettent de catégoriser, selon l'ordre dominant, la géométrie locale des pixels en 4 catégories (6 en tenant compte de la polarité) :

| 0 | 1 | 2 | |
|---|---|---|---|
| $\|\|\nabla_I\|\| \simeq 0$<br>$\|\|H_I\|\|_F \simeq 0$ | $\|\|\nabla_I\|\| \gg 0$<br>$\|\|H_I\|\|_F \simeq 0$ | $\|\|H_I\|\|_F \gg 0$ | |
| | | $\Lambda_I \lambda_I > 0$ | $\Lambda_I \lambda_I < 0$ |
| Plateau | Contour | Courbure elliptique | Courbure tubulaire |
|  |  |  |    |
| | | $\Lambda_I < 0$<br>$\lambda_I < 0$ | $\Lambda_I > 0$<br>$\lambda_I > 0$ |
| | | | $\Lambda_I < 0$<br>$\lambda_I > 0$ |
| | | | $\Lambda_I > 0$<br>$\lambda_I < 0$ |

# ESTIMATION DES DERIVEES ET ESPACE D'ECHELLES

L'idée clef des espace d'échelles pour le traitement d'images est que toute mesure est relative à une échelle d'estimation.

En particulier une dérivée n'a de sens qu'estimée à une échelle donnée, correspondant à une hypothèse de régularité qui est explicitement réalisée par lissage de l'image. Cette estimation repose sur la commutativité entre dérivation et convolution :

$$\partial^n (I \star g) = I \star (\partial^n g)$$

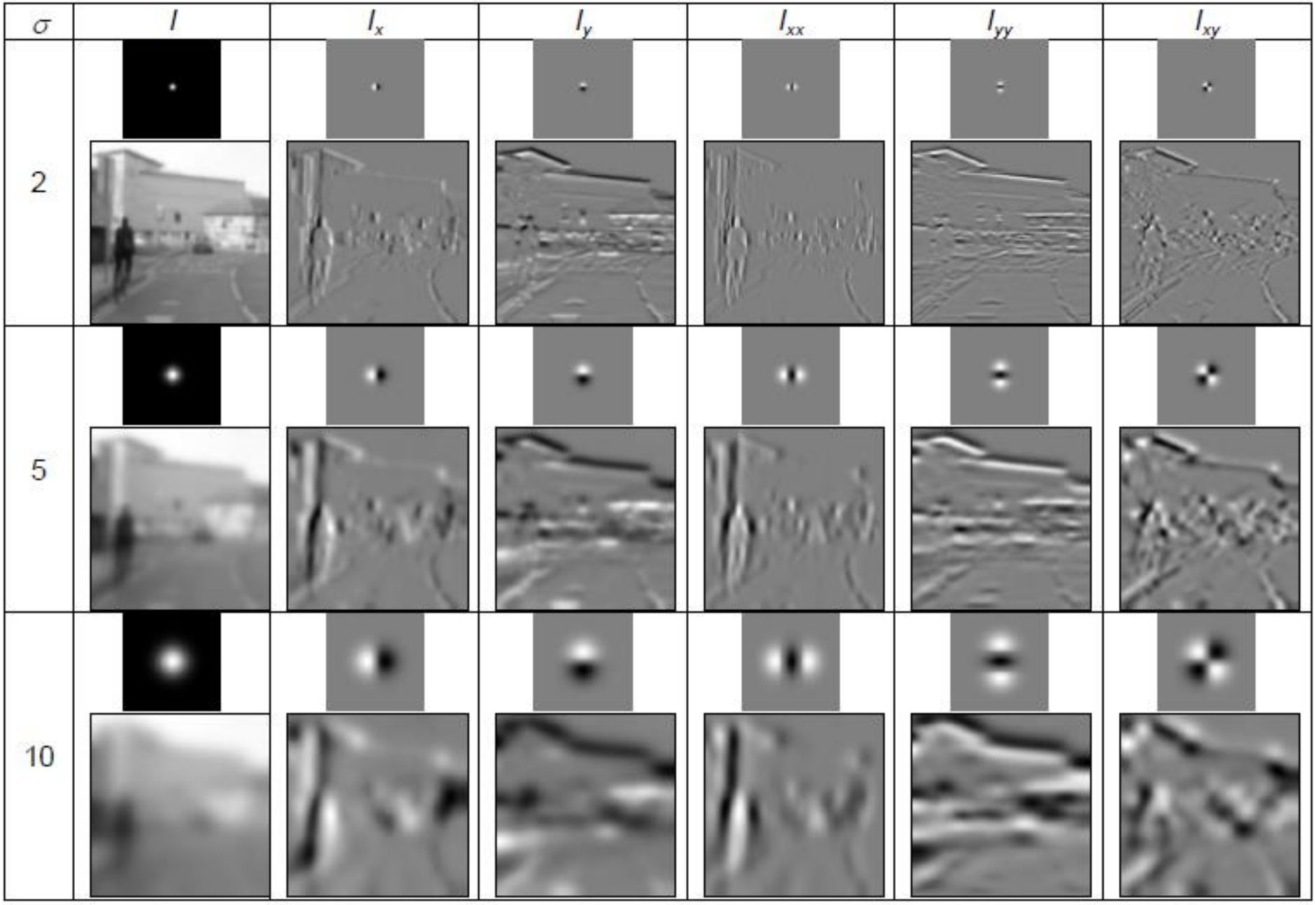
Dans le cadre des espaces d'échelles gaussien, le noyau de convolution *g* s'identifie au noyau gaussien 2d d'écart-type σ :

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

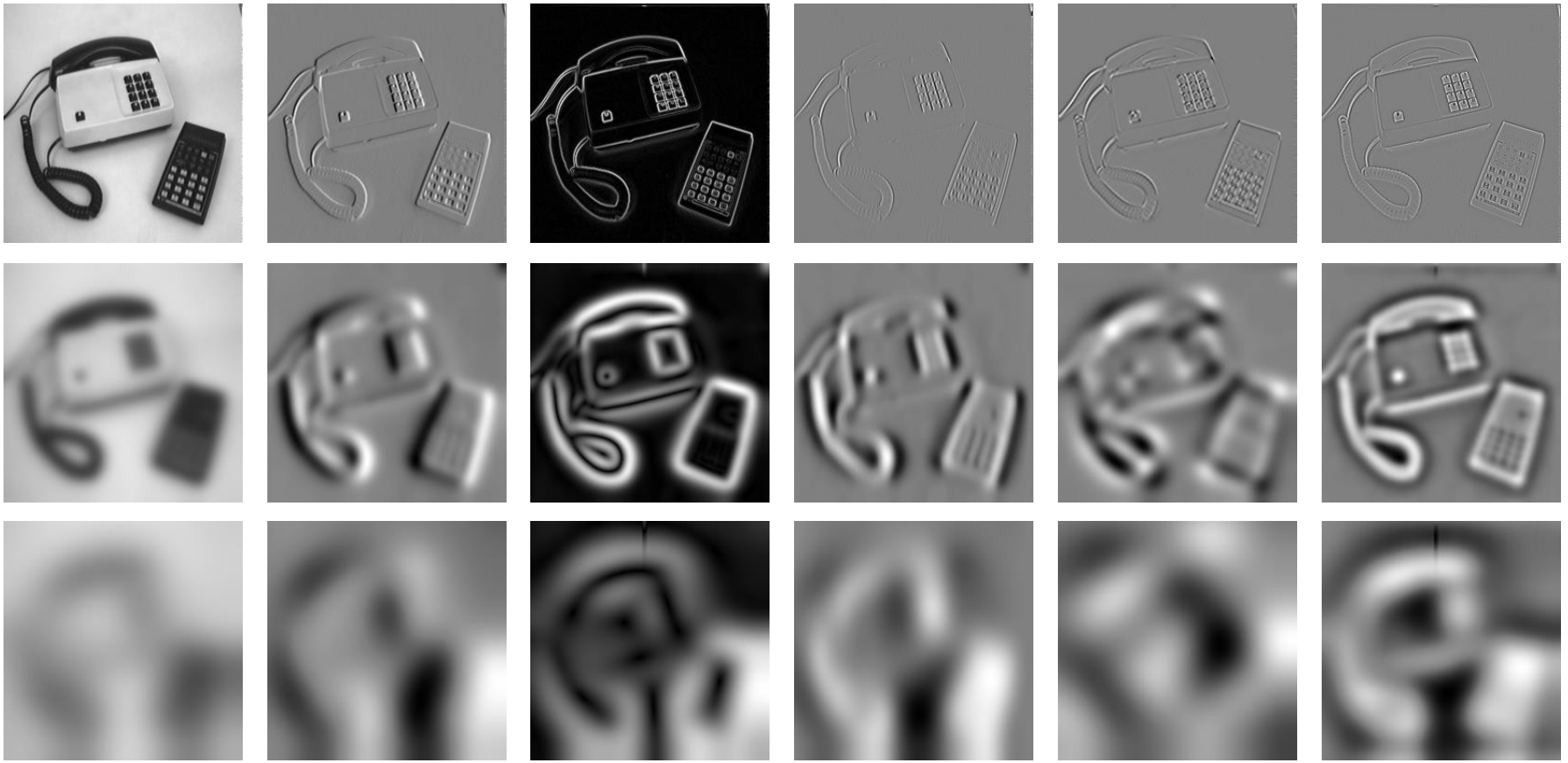
Les dérivées de l'image *I* estimées à l'échelle σ sont donc définies par les convolutions avec la dérivée de gaussienne correspondante :

$$\left(\frac{\partial^{i+j} I}{\partial x^i \partial y^j}\right)_\sigma \overset{\textbf{déf.}}{=} I \star \left(\frac{\partial^{i+j} G_\sigma}{\partial x^i \partial y^j}\right)$$

# DERIVEES MULTI-ECHELLES ET NOYAUX DE CONVOLUTION ASSOCIES



# ESPACE D'ECHELLES GAUSSIEN ET GRANDEURS DIFFERENTIELLES



*échelle*

$I$ $\quad \frac{\partial I}{\partial x} \quad \|\nabla I\| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad \frac{\partial^2 I}{\partial x^2} \quad \frac{\partial^2 I}{\partial x \partial y} \quad \Delta I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$

# PRIMITIVES VISUELLES POUR L'INDEXATION ET LA RECONNAISSANCE

Le niveau de la représentation, du strictement local au complètement global, forme une caractéristique fondamentale d'une primitive visuelle.

**Local : plutôt géométrique (direction, courbure,…)**

**Global : plutôt statistique (histogramme, spectre de fréquences,…)**

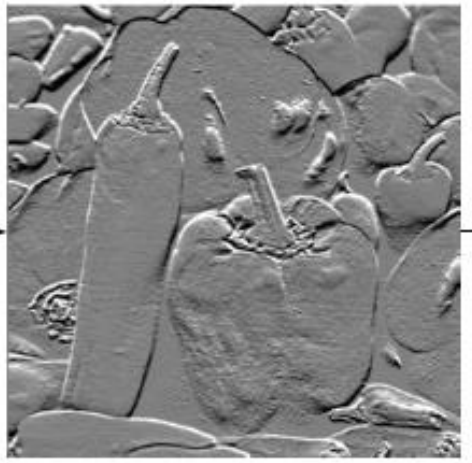
Les espaces d'échelles permettent de réaliser un continuum du local vers le global.

Dans la suite :

- Détection de contours (Zéros du laplacien)
- Calcul des points anguleux (Harris)
- Calcul des blobs (SIFT)
- Calcul des descripteurs (invariants différentiels).

# CONTOURS : METHODES BASIQUES

*Original*



*Sobel vertical :* $\frac{\partial I}{\partial y}$

*Sobel horizontal :* $\frac{\partial I}{\partial x}$

|...|

|...|

+

*Norme $L_1$ du gradient de Sobel*

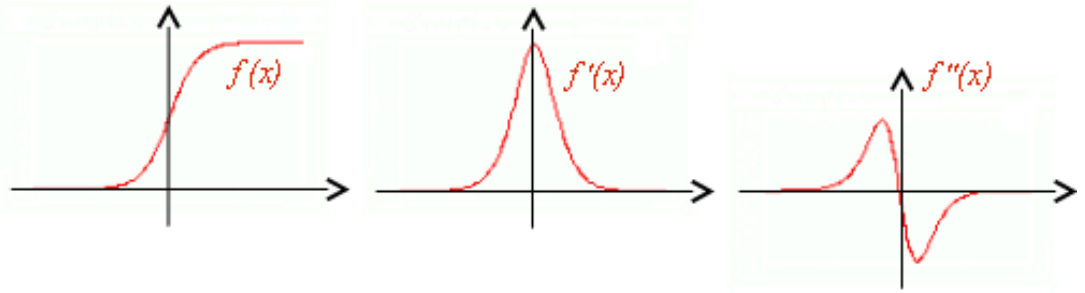
*Seuillage*

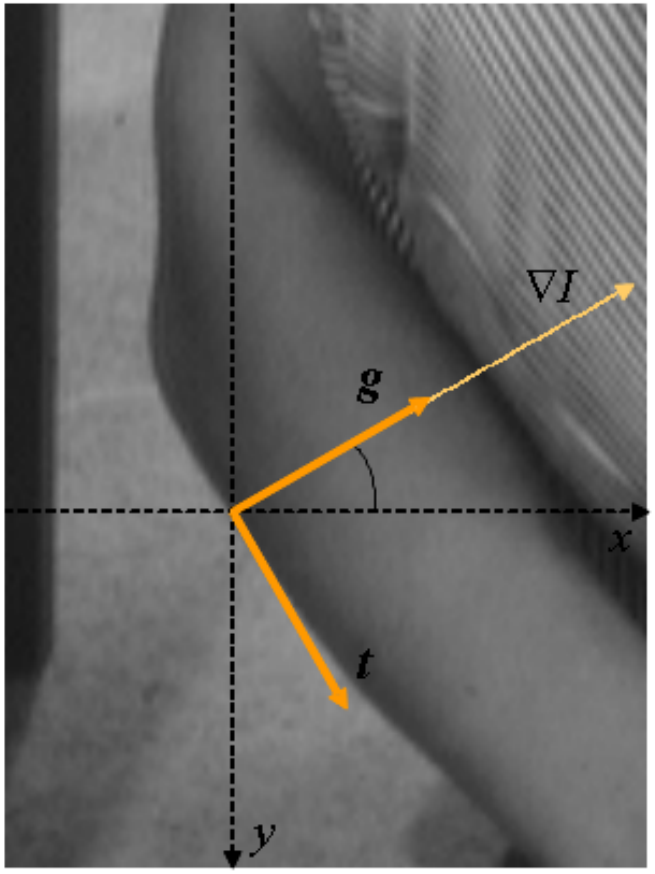
*Contours épais*

- *Amincissement*
- *Poursuite*
- *Fermeture*

# CONTOURS : METHODE ANALYTIQUE

En 1d, un contour correspond à un maximum de la dérivée première, c'est-à-dire à un passage par zéro de la dérivée seconde :



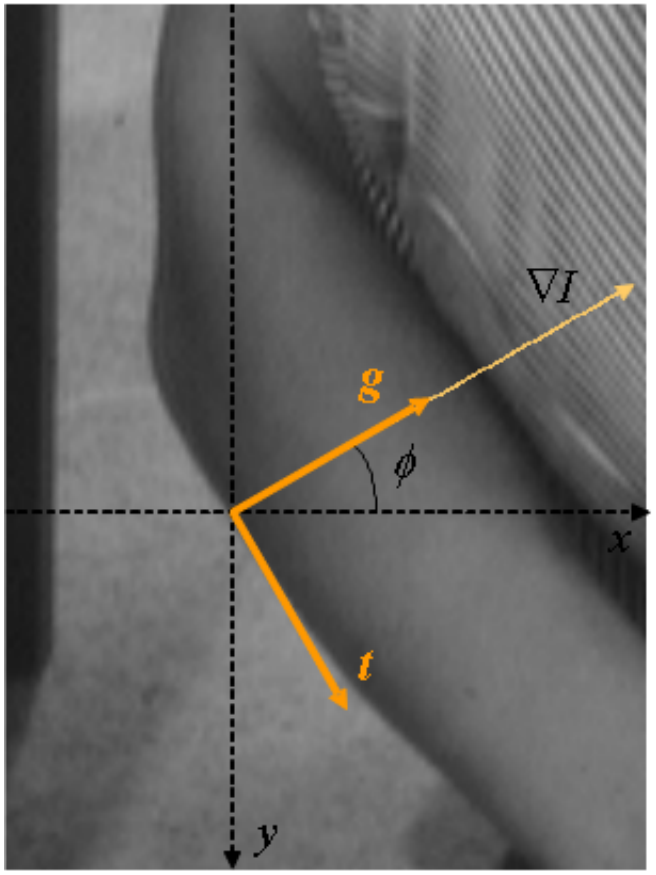




Pour retrouver cette propriété en 2d, il faut se placer dans la direction du gradient. Les contours sont donc définis comme les maxima de la dérivée première dans la direction du gradient, c'est-à-dire des passages par zéros de la dérivée seconde dans la direction du gradient, soit :

$$I_{gg}=0$$

$$\left(I_{gg}=\frac{\partial^2 I}{\partial g^2}\right)$$

# CONTOURS : METHODE ANALYTIQUE




Exprimons la fonction image et ses dérivées dans le repère local au gradient $(g,t)$ :

$$\begin{cases} g = x\cos\phi + y\sin\phi \\ t = -x\sin\phi + y\cos\phi \end{cases} \qquad \begin{cases} x = g\cos\phi + t\sin\phi \\ y = g\sin\phi - t\cos\phi \end{cases}$$

$$\text{avec} \quad \phi = arg(\nabla I) = \arctan\left(\frac{I_x}{I_y}\right)$$

Calcul des dérivées premières :

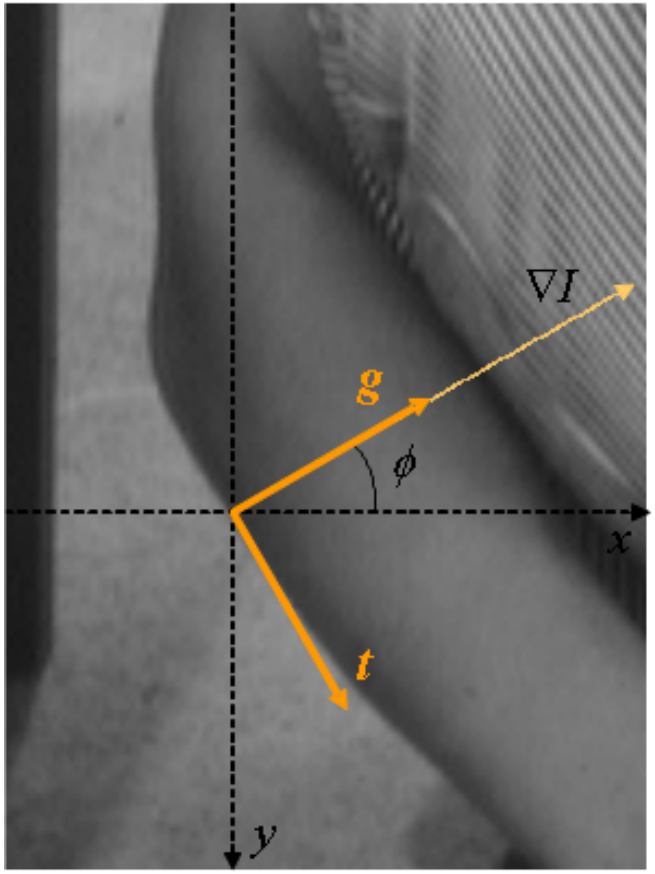
$$\begin{cases} I_g = I_x\cos\phi + I_y\sin\phi \\ I_t = I_x\sin\phi - I_y\cos\phi \end{cases}$$

Et puisque $\cos\phi = \frac{I_x}{\|\nabla I\|}$ et $\sin\phi = \frac{I_y}{\|\nabla I\|}$ on a :

$$\begin{cases} I_g = \|\nabla I\| \\ I_t = 0 \end{cases}$$

- La composante tangentielle $t$ correspond à la direction de l'isophote ou ligne de niveau.
- La composante gradient $g$ correspond à la direction principale de variation.

# CONTOURS : METHODE ANALYTIQUE




Calcul des dérivées secondes :

$$\begin{cases} I_{gg} = I_{xx}\cos^2\phi + 2I_{xy}\cos\phi\sin\phi + I_{yy}\sin^2\phi \\ I_{tt} = I_{xx}\sin^2\phi - 2I_{xy}\cos\phi\sin\phi + I_{yy}\cos^2\phi \end{cases}$$

Et finalement, en remplaçant cos ϕ et sin ϕ par leur expression en fonction du gradient, l'équation des contours devient :

$$I_{xx}I_x^2 + 2I_{xy}I_xI_y + I_{yy}I_y^2 = 0$$

Soit 5 dérivées à calculer. On peut néanmoins gagner beaucoup en approximant la dérivée seconde dans la direction de $g$ par le laplacien. En effet, on peut remarquer que :

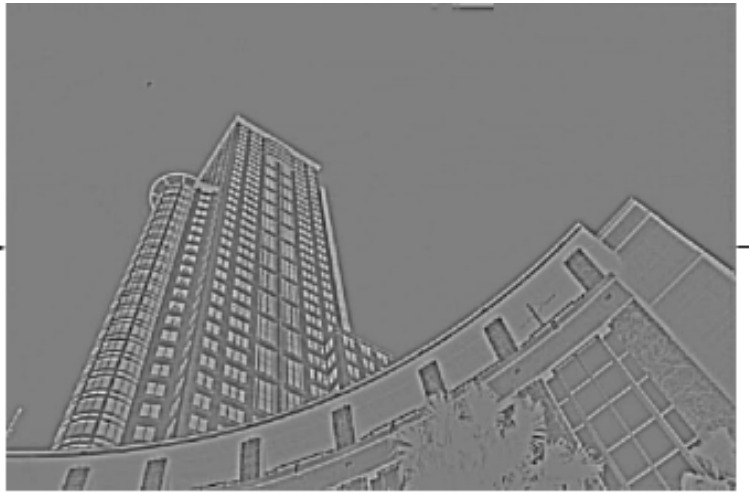
$$I_{gg} + I_{tt} = I_{xx} + I_{yy} = \Delta I$$

Le laplacien est donc invariant par rotation.

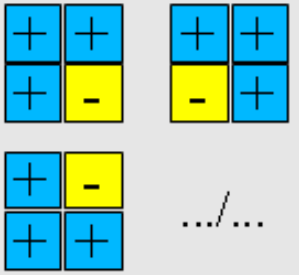
Or $I_{tt}$, la dérivée seconde dans la direction de $t$, correspond à la courbure de la ligne isophote (ligne de niveau). *Si cette courbure est faible*, on a : $I_{tt} \simeq 0 \Rightarrow \Delta I \simeq I_{gg}$

Finalement, l'équation des contours devient : $\Delta I = 0$

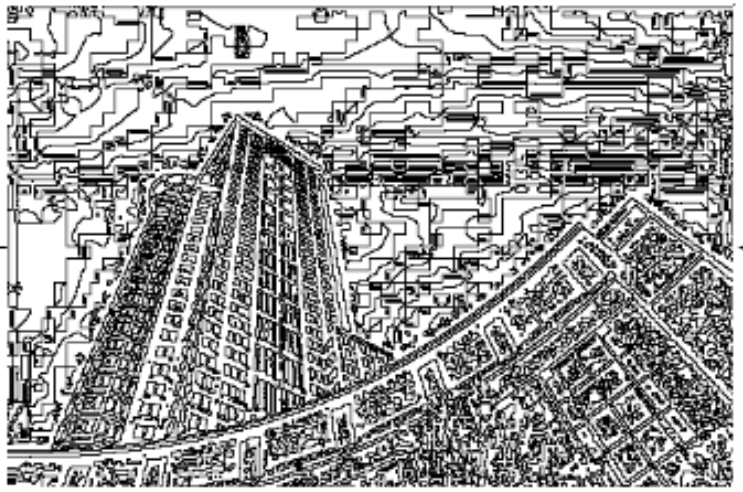
# CONTOURS : PASSAGES PAR ZERO DU LAPLACIEN



*Laplacien*



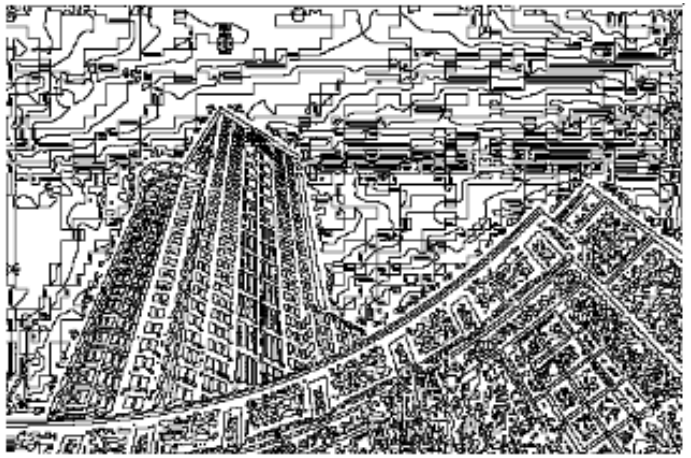

Détection des changements de polarité dans un voisinage 2x2



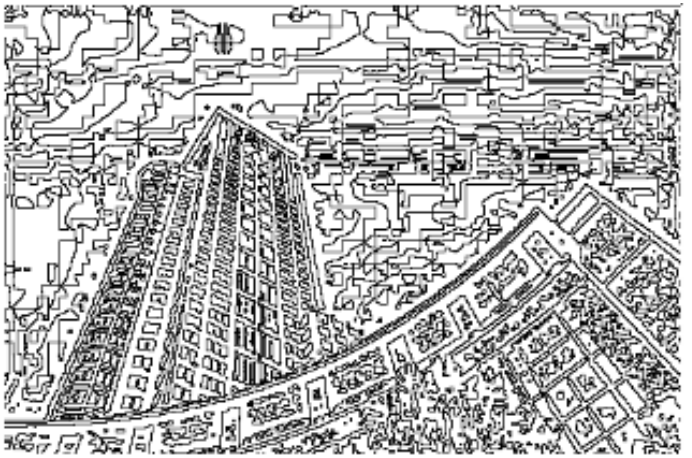
*Passages par zéro*

- Seuiller les passages par zéro selon le contraste.
- Sélectionner les structures en fonction de l'échelle

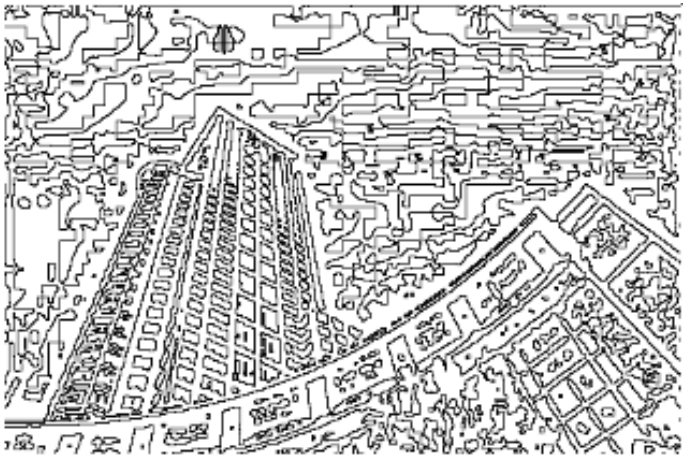
# CONTOURS MULTI ECHELLES



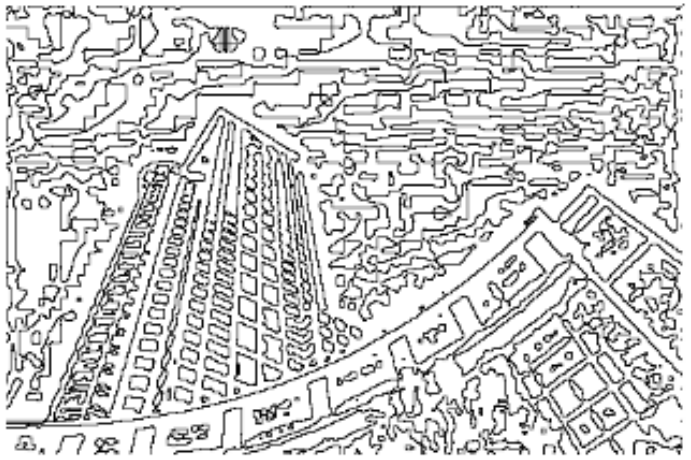
$\sigma = 1.0$



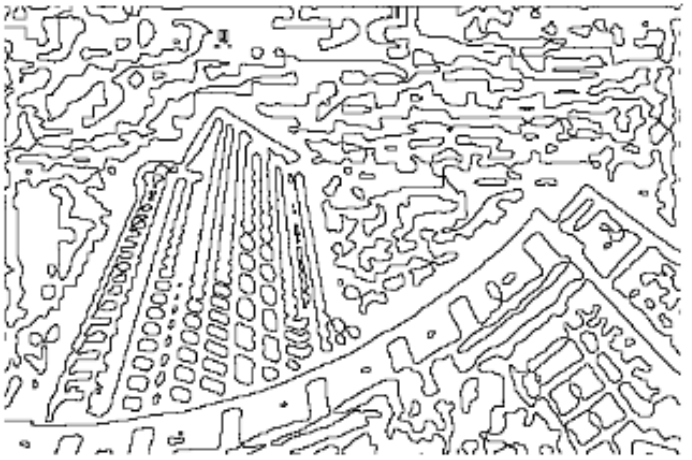
$\sigma = 1.5$



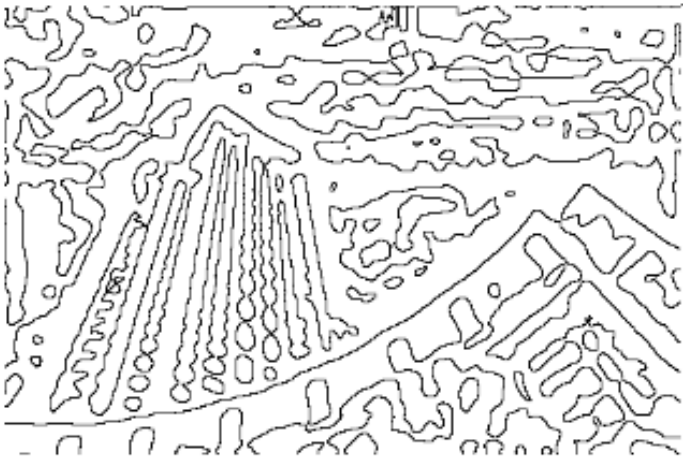
$\sigma = 2.0$



$\sigma = 2.5$

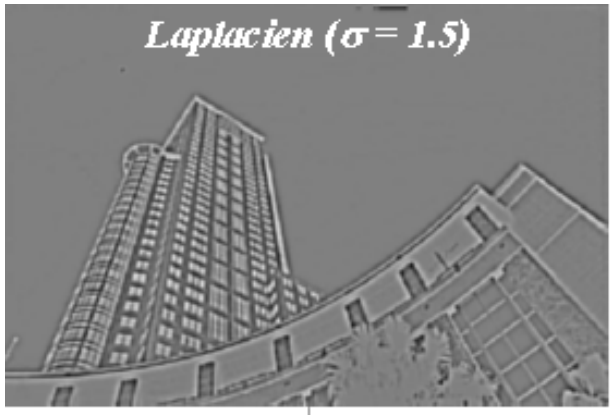


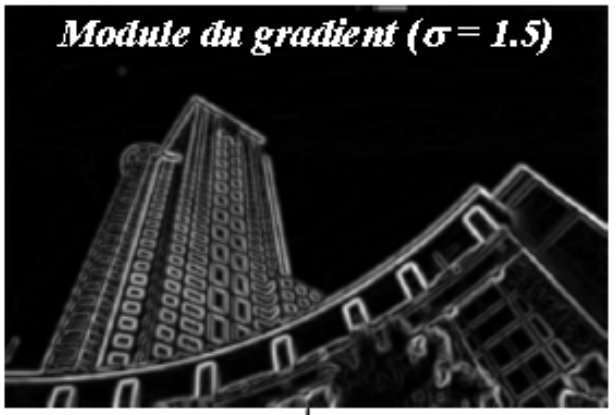
$\sigma = 3.5$



$\sigma = 5.0$

# CONTOURS ET CONTRASTE

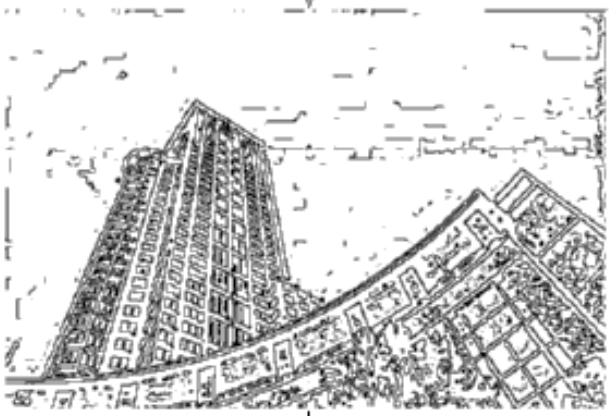



passages par zéro

Seuil bas (s = 0.5)

Seuil haut (s = 8.0)

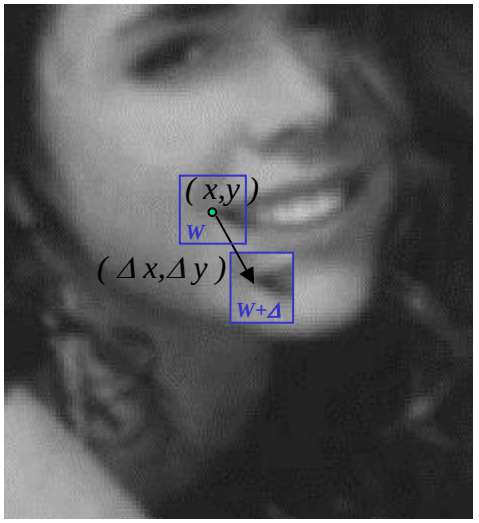


Seuil par hystérésis

Le seuil par hystérésis consiste à sélectionner l'ensemble des composantes connexes dans l'image de seuil bas, qui ont au moins 1 pixel dans l'image de seuil haut (voir reconstruction géodésique).

# POINTS ANGULEUX ET MATRICE D'AUTOCORRELATION

Les points anguleux (ou points d'intérêt, points saillants,...) sont des points « qui contiennent beaucoup d'information » relativement à l'image. Ce sont des points aux voisinages desquels l'image *varie significativement dans plusieurs directions*.




Une mesure des variations locales de l'image $I$ au point $(x,y)$ associée à un déplacement $(\Delta x,\Delta y)$ est fournie par la *fonction d'autocorrélation* :

$$\chi(x,y) = \sum_{(x_k,y_k)\in W} \left(I(x_k,y_k) - I(x_k+\Delta x, y_k+\Delta y)\right)^2$$

Où $W$ est une fenêtre centrée au point $(x,y)$.

Or, en utilisant une approximation du premier ordre :

$$I(x_k+\Delta x, y_k+\Delta y) \approx I(x_k,y_k) + \left(\frac{\partial I}{\partial x}(x_k,y_k) \quad \frac{\partial I}{\partial y}(x_k,y_k)\right)\cdot\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}$$

Et donc :

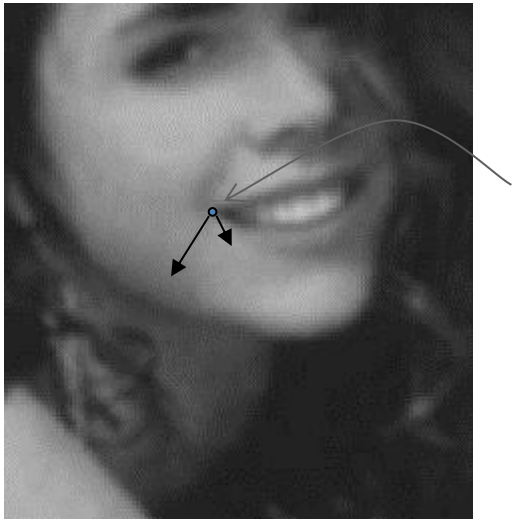
$$\chi(x,y) = \sum_{(x_k,y_k)\in W}\left(\left(\frac{\partial I}{\partial x}(x_k,y_k) \quad \frac{\partial I}{\partial y}(x_k,y_k)\right)\cdot\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}\right)^2 = (\Delta x \quad \Delta y)\underbrace{\begin{pmatrix}\sum_{(x_k,y_k)\in W}\left(\frac{\partial I}{\partial x}(x_k,y_k)\right)^2 & \sum_{(x_k,y_k)\in W}\frac{\partial I}{\partial x}(x_k,y_k)\cdot\frac{\partial I}{\partial y}(x_k,y_k)\\ \sum_{(x_k,y_k)\in W}\frac{\partial I}{\partial x}(x_k,y_k)\cdot\frac{\partial I}{\partial y}(x_k,y_k) & \sum_{(x_k,y_k)\in W}\left(\frac{\partial I}{\partial y}(x_k,y_k)\right)^2\end{pmatrix}}_{\Xi(x,y)}\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}$$

*Matrice d'autocorrélation* de l'image $I$ en $(x,y)$

# MATRICE D'AUTOCORRELATION ET DETECTEUR DE HARRIS

$$\Xi(x,y) = \begin{pmatrix} \sum_{(x_k,y_k)\in W} \left(\frac{\partial I}{\partial x}(x_k,y_k)\right)^2 & \sum_{(x_k,y_k)\in W} \frac{\partial I}{\partial x}(x_k,y_k)\cdot\frac{\partial I}{\partial y}(x_k,y_k) \\ \sum_{(x_k,y_k)\in W} \frac{\partial I}{\partial x}(x_k,y_k)\cdot\frac{\partial I}{\partial y}(x_k,y_k) & \sum_{(x_k,y_k)\in W} \left(\frac{\partial I}{\partial y}(x_k,y_k)\right)^2 \end{pmatrix}$$

La matrice d'autocorrélation $\Xi$ représente la variation locale de l'image $I$ en $(x,y)$. $(x,y)$ sera considéré comme un point anguleux de I si pour tous les déplacements $(\Delta x,\Delta y)$, la quantité $(\Delta x,\Delta y).\Xi(x,y).(\Delta x,\Delta y)^t$ est grande.



Les points anguleux sont les points $(x,y)$ pour lesquels la matrice d'autocorrélation $\Xi(x,y)$ a *deux valeurs propres grandes*.

Cela correspond aux points pour lesquels il existe localement une base de vecteurs propres de $\Xi$ décrivant des variations locales importantes de l'image.

Le *détecteur de Harris* calcule une *fonction d'intérêt* $\Theta(x,y)$ :

$$\Theta(x,y) = \det \Xi - \alpha \,\mathrm{trace}\, \Xi$$

Le premier terme correspond au produit des valeurs propres, le second terme pénalise les points de contours avec une seule forte valeur propre.

Les points d'intérêt correspondent aux maxima locaux de la fonction $\Theta$ qui sont au delà d'un certain seuil (typiquement 1% de la valeur max de $\Theta$).

***[Harris 88]***

# CALCUL DE LA FONCTION D'INTERET DE HARRIS

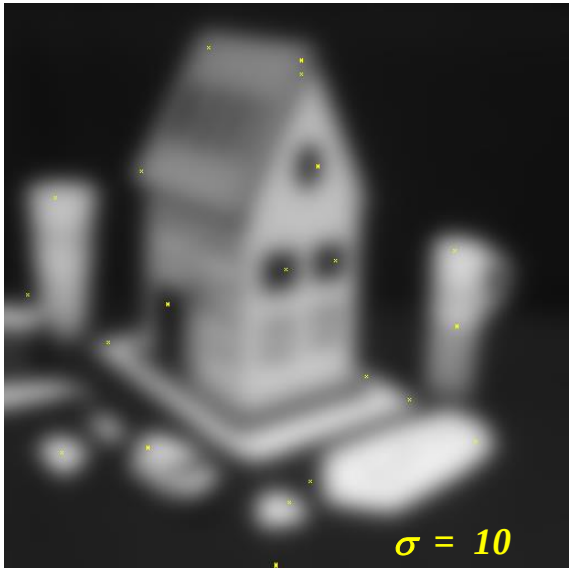
1. On calcule les dérivées premières à partir des dérivées de gaussienne (écart-type $\sigma_1$)
2. On calcule les termes de la matrice d'autocorrélation $\Xi$ en calculant une moyenne locale des dérivées sous la forme d'une gaussienne (écart-type $\sigma_2$, typiquement $\sigma_2 = 2\ \sigma_1$)
3. On calcule la fonction d'intérêt : $\Theta = \det(\Xi) - \alpha\ \mathrm{trace}(\Xi)$ (typiquement $\alpha = 0{,}06$).
4. On calcule les maxima locaux de $\Theta$ supérieurs à un certain seuil (typiquement 1% de $\Theta_{max}$).


***$\sigma_1$ est donc le paramètre d'échelle du détecteur de Harris, qui détermine la portée des opérations de dérivations et d'intégrations (moyennes).***

# POINTS DE HARRIS MULTI-ECHELLES




*Points de Harris obtenus en calculant les dérivées premières par convolution avec une dérivée de gaussienne d'écart-type σ.*

# DETECTEUR SIFT : EXTREMA DANS L'ESPACE D'ECHELLES

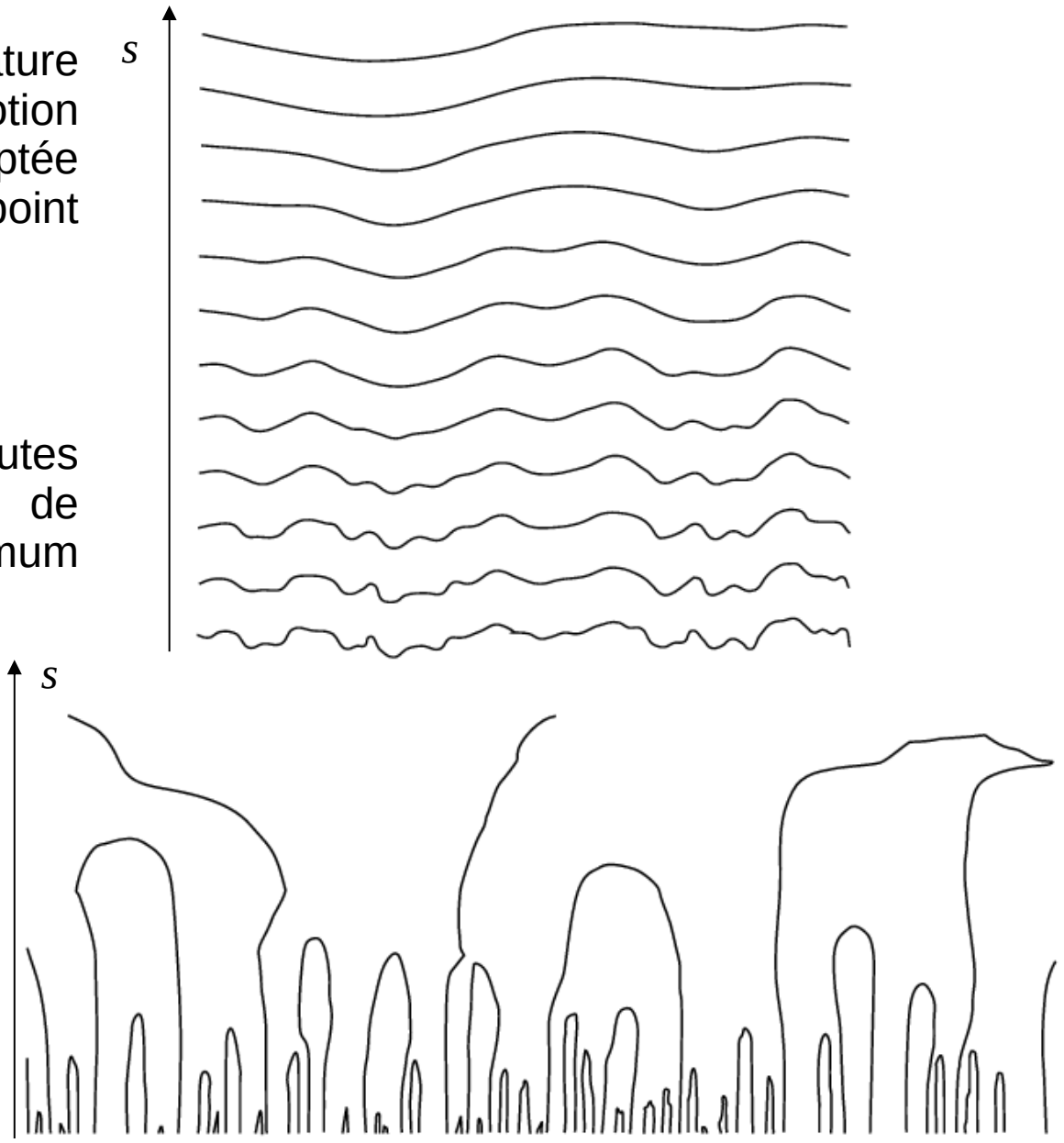
Le détecteur SIFT (Scale Invariant Feature Transform) utilise une approche de la notion de structure / point d'intérêt mieux adaptée aux grandes échelles que celle de point anguleux :

Le *blob* (structure elliptique)

Cette structure peut se caractériser à toutes les échelles et correspond au point de l'espace-échelle ($x,y,s$) où un extremum local disparaît.

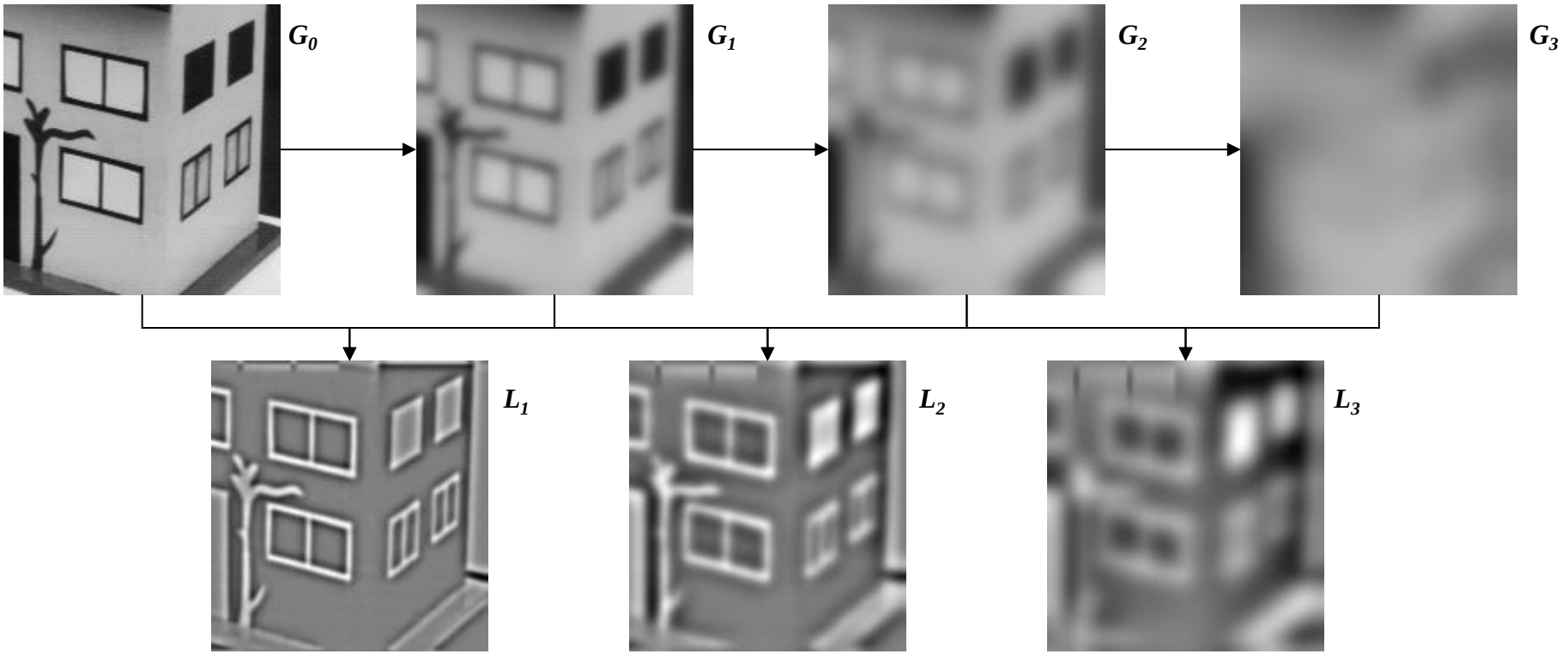
Cf principe de causalité des espaces d'échelles.

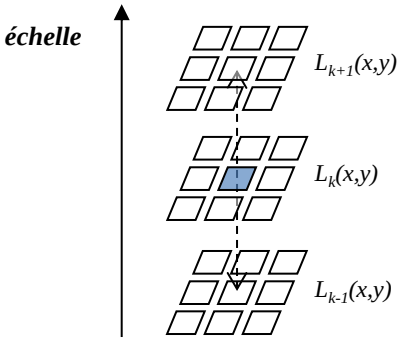
En 1d (ci-contre) : point d'échelle s maximum sur chaque courbe de l'empreinte (fingerprint) de l'espace d'échelles.




*Espace d'échelle gaussien d'un signal 1d (en haut) et position des extrema dans l'espace-échelle (x,s) (en bas).*

# DETECTEUR SIFT : EXTREMA DANS L'ESPACE D'ECHELLES




La fonction $G_k(x,y) = G(x,y,k\sigma)$ est l'image convoluée par une gaussienne d'écart-type $k\sigma$. Les fonctions $L_k(x,y)$ correspondent à la différence (ici normalisée) entre 2 gaussiennes adjacentes.

La fonction $L_k(x,y)$ est une représentation laplacienne de l'image, qui correspond à une décomposition fréquentielle localisée : contribution des structures contrastées d'échelle (de « taille ») $k\sigma$ au point $(x,y)$.

Les points sélectionnés par SIFT sont les maxima et les minima locaux de la fonction $L_k(x,y)$, à la fois dans l'échelle courante et dans les échelles adjacentes (voir ci-contre).

***[Lowe 04]***

# POINTS D'INTERET SIFT

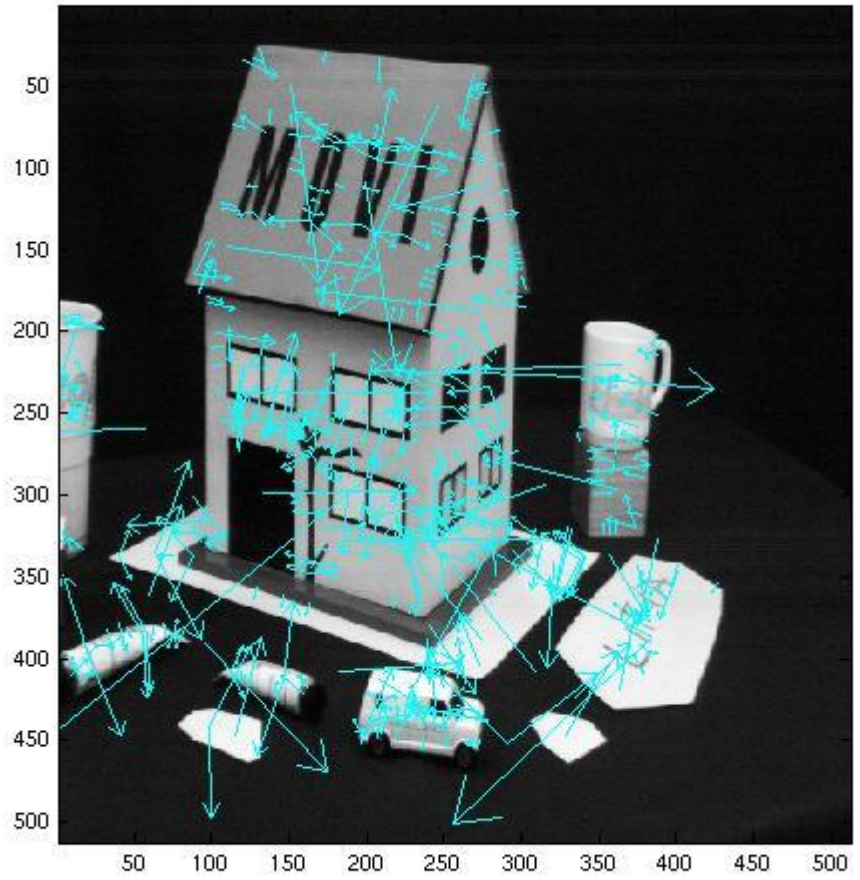



***Image 1 : 589 points détectés.***

Pour chaque extrema de l'espace d'échelle des différences de gaussiennes (point d'intérêt SIFT), on calcule la direction associée par :

$$\theta(x,y) = \arctan\left(\frac{G_y^\sigma(x,y)}{G_x^\sigma(x,y)}\right)$$

avec $G_x^\sigma(x,y) = \frac{\partial}{\partial x} G(x,y,\sigma) = I(x,y) * \frac{\partial}{\partial x} g_\sigma(x,y)$

(où $\sigma$ est l'échelle sélectionnée)

Ci-contre, point d'intérêt SIFT : la direction de la flèche représente la direction $\theta$ et sa longueur l'échelle $\sigma$ associée.

***[Lowe 04]***

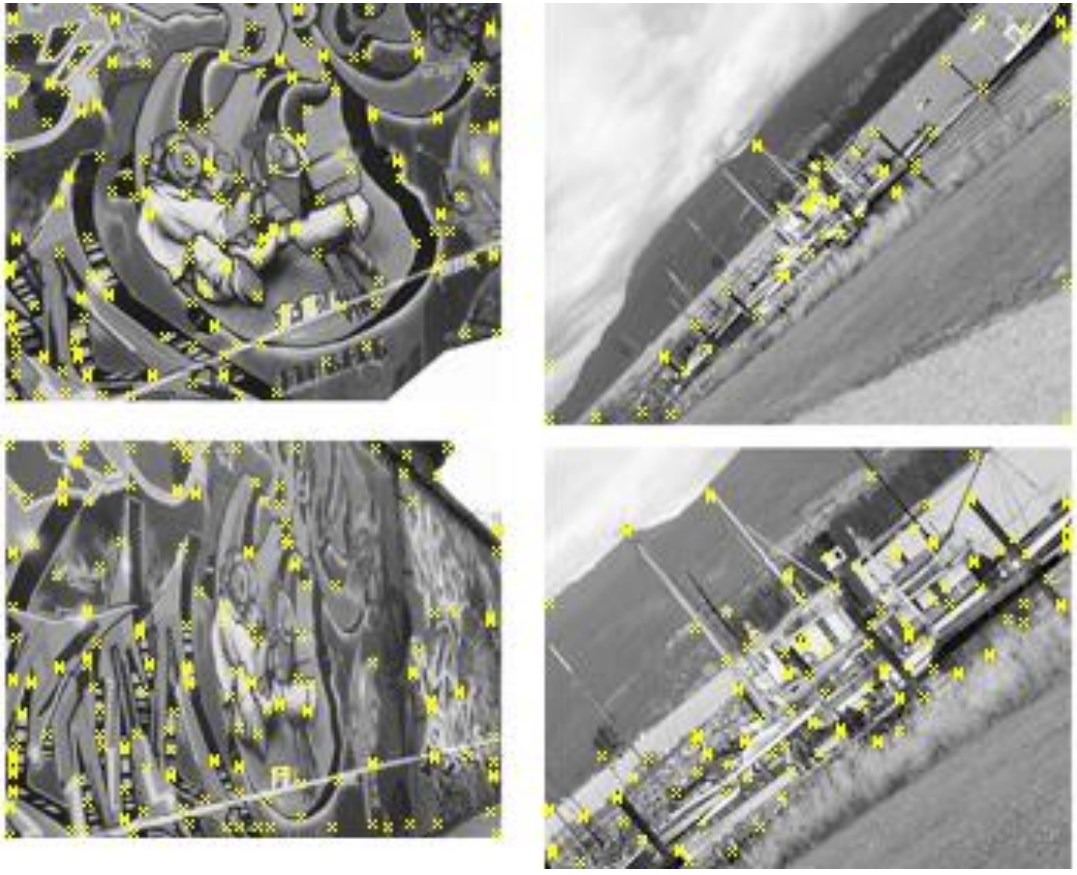
# EVALUATION DES DETECTEURS DE POINTS D'INTERET

La plupart des détecteurs de point d'intérêt sont définis indépendamment des descripteurs avec lesquels on les utilise. Il est donc nécessaire de pouvoir les évaluer en eux-mêmes.

Les propriétés recherchées d'un bon détecteur :

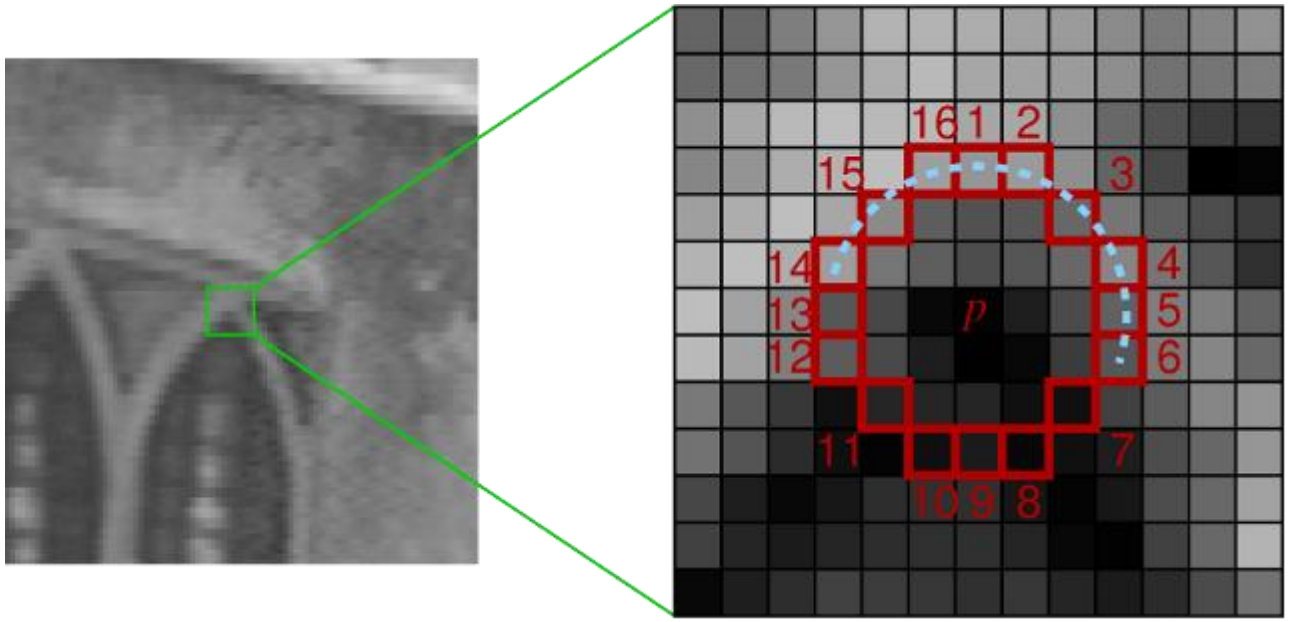
- **Répétabilité** : le point doit apparaître aux mêmes endroits quelque soit la déformation.
- **Représentativité** : les points doivent être le plus nombreux possible.
- **Efficacité** : le détecteur doit être rapide à calculer (cf SURF, FAST)

(Rq : répétabilité et représentativité ne sont pas indépendants !)



*[Schmid 2000]*

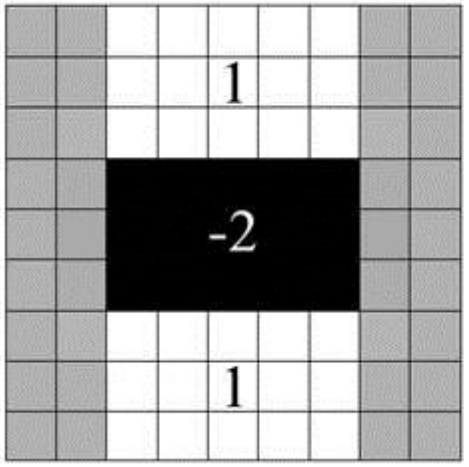
# DES DETECTEURS RAPIDES A CALCULER ?

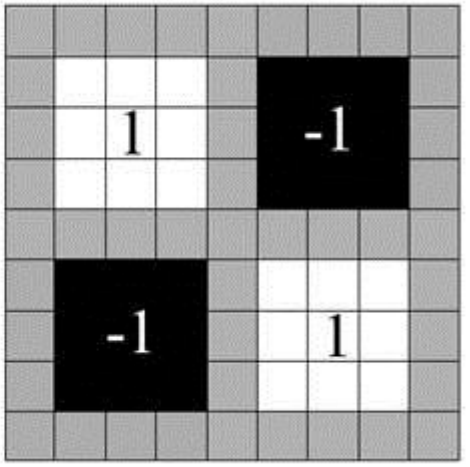



Le détecteur FAST sélectionne les points dont le voisinage circulaire présente des plages contiguës assez longues de points significativement plus clairs (resp. plus sombres).

*[Rosten 05]*

Le détecteur SURF approxime le calcul des dérivées secondes dans des voisinages rectangulaires à l'aide d'images intégrales, et sélectionne les maxima locaux du déterminant de la matrice hessienne.




*[Bay 06]*

# DESCRIPTEURS : INVARIANTS DIFFERENTIELS

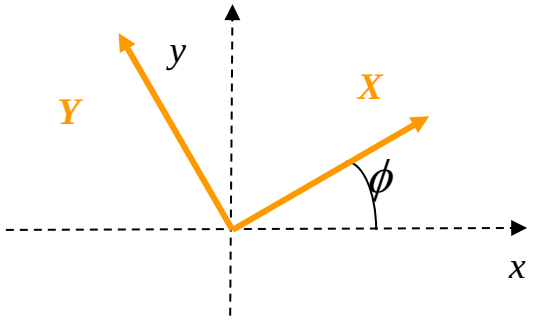
***Objectif*** : *représenter* les points d'intérêt par des *indices* qui soient *invariants* par *rotation* et par *changement d'échelle*.

Le principe utilisé ici est basé sur l'utilisation des dérivées spatiales multi-échelle :

Le « *jet local* » de I : $L_{ij}^{\sigma} = I * G_{ij}^{\sigma}$ avec : $G_{ij}^{\sigma} = \frac{\partial^{i+j}}{\partial x^i \partial y^j} G^{\sigma}$ et : $G^{\sigma}(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$

On notera : $\left\{L_{ij}^{\sigma}; 0 \le i+j \le 3\right\} = \left\{L, L_x, L_y, L_{xx}, L_{xy}, L_{yy}, L_{xxx}, L_{xxy}, L_{xyy}, L_{yyy}\right\}$ (dérivées jusqu'au 3e ordre)

L'idée est de *combiner ces dérivées* pour obtenir des grandeurs *invariantes par rotation* :




Par exemple, le laplacien $I_{xx} + I_{yy}$ est invariant par rotation :

$$\begin{cases} x = X\cos\phi + Y\sin\phi \\ y = X\sin\phi - Y\cos\phi \end{cases} \qquad \begin{cases} X = x\cos\phi + y\sin\phi \\ Y = -x\sin\phi + y\cos\phi \end{cases}$$

$$\begin{cases} I_X = I_x\cos\phi + I_y\sin\phi \\ I_Y = I_x\sin\phi - I_y\cos\phi \end{cases} \quad \text{et :} \quad \begin{cases} I_{XX} = I_{xx}\cos^2\phi + 2I_{xy}\cos\phi\sin\phi + I_{yy}\sin^2\phi \\ I_{YY} = I_{xx}\sin^2\phi - 2I_{xy}\cos\phi\sin\phi + I_{yy}\cos^2\phi \end{cases}$$

et donc : $I_{XX} + I_{YY} = I_{xx} + I_{yy}$

# DESCRIPTEURS : INVARIANTS DIFFERENTIELS

On peut ainsi construire toute une famille de grandeurs invariantes par rotation : les *invariants différentiels de Hilbert*.

$$\Psi = \begin{pmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ij} \\ \varepsilon_{ij}\left(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l\right) \\ L_{ijj} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{pmatrix}$$

Avec :
$$\varepsilon_{xx} = \varepsilon_{yy} = 0$$
$$\varepsilon_{xy} = -\varepsilon_{yx} = 1$$

(notations d'Einstein : sommations sur les indices), par ex :

$$\Psi_2 = L_i L_{ij} L_j = L_x L_{xx} L_x + 2 L_x L_{xy} L_y + L_y L_{yy} L_y$$
$$\Psi_7 = -\varepsilon_{ij} L_{jkl} L_i L_k L_l = L_{xxy}\left(-L_x L_x L_x + 2 L_x L_y L_y\right)$$
$$+ L_{xyy}\left(-2 L_x L_x L_y + L_y L_y L_y\right) - L_{yyy} L_x L_y L_y + L_{xxx} L_x L_x L_y$$

NB : invariance par rotation du noyau gaussien !

Les vecteurs $\Psi$ sont donc calculés pour tous les points d'intérêt à différentes échelles, et appariés en utilisant une distance (e.g. distance euclidienne).

***[Schmid et Mohr 97]***

# DESCRIPTEURS SIFT : HISTOGRAMMES LOCAUX D'ORIENTATION

Les descripteurs associés aux points d'intérêt SIFT sont des histogrammes des orientations locales autour du point d'intérêt.

• On divise l'espace autour de chaque point d'intérêt $(x,y)$ en $N^2$ carrés 4x4.

• On calcule le gradient $(G_x(a,b,\sigma), G_y(a,b,\sigma))$ pour les 4x4x$N^2$ points $(a,b)$.

• Pour chaque carré 4x4, on calcule un histogramme des orientations quantifiées en 8 directions, en pondérant par : (1) le module du gradient (2) l'inverse de la distance au point d'intérêt $(x,y)$.

• Pour être invariant en rotation : l'orientation locale du point d'intérêt $\theta(x,y)$ est utilisée comme *origine* (orientation nulle) des histogrammes.

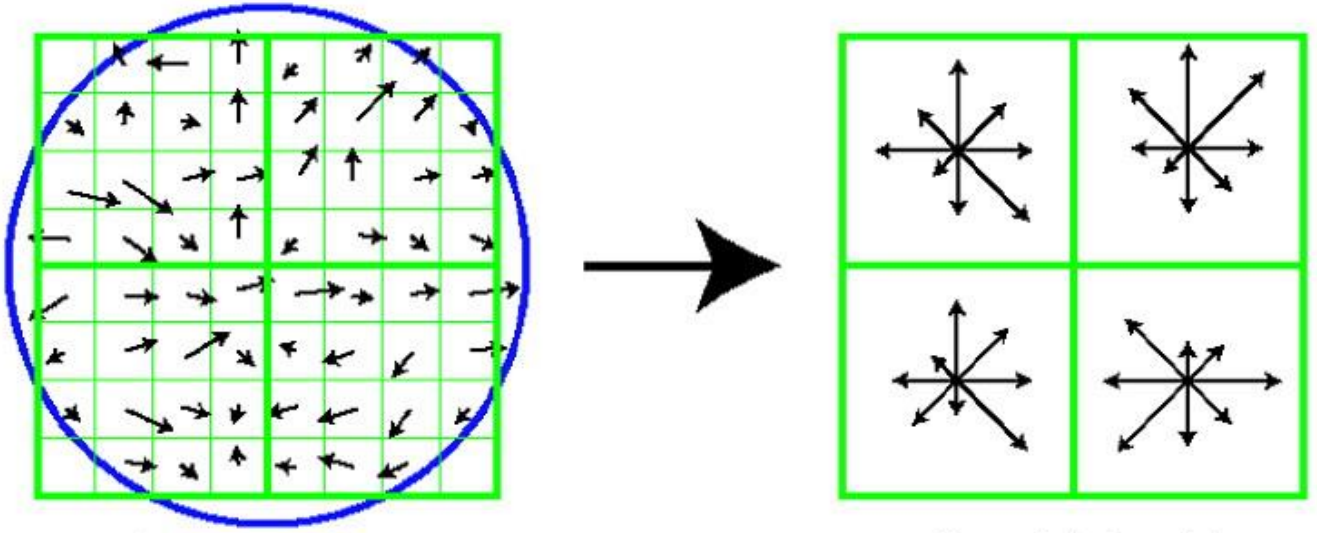


Image gradients

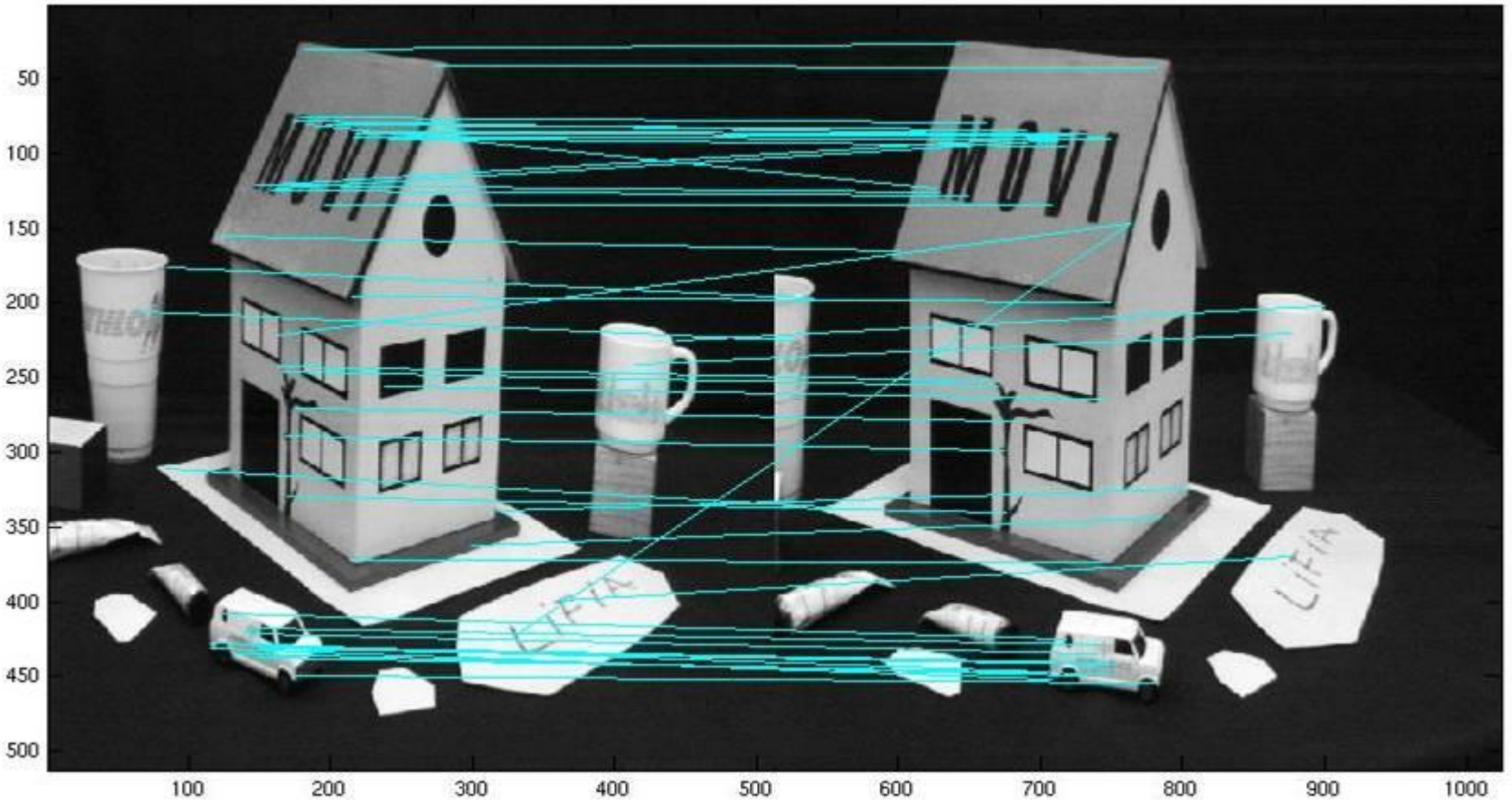
ex : $N = 2$

Keypoint descriptor

Les descripteurs formés sont donc des vecteurs de taille 8x$N^2$, qui seront appariés en utilisant une distance (e.g. distance euclidienne)

***[Lowe 04]***

# APPARIEMENT DES POINTS SIFT




***Résultat d'appariement par SIFT entre l'image (2) à gauche, 510 points détectés, et l'image (1) à droite, 589 points détectés. 51 points ont été appariés, ce qui correspond à une distance euclidienne entre les descripteurs en deçà d'un certain seuil.***

***[Lowe 04]***

# APPARIEMENT DE PRIMITIVES : METRIQUES

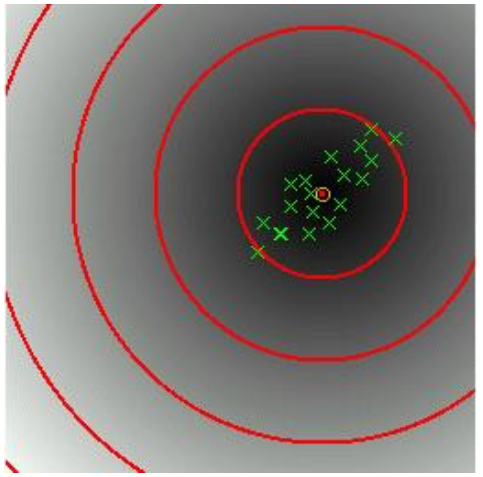
L'appariement entre primitives repose donc en général sur la comparaison de descripteurs locaux deux à deux.

Si le couple détecteur / descripteur possède l'invariance voulue, la comparaison peut être réalisée par une métrique simple :

La distance euclidienne :

$$\delta_e(x, x')^2 = (x - x')^T (x - x')$$

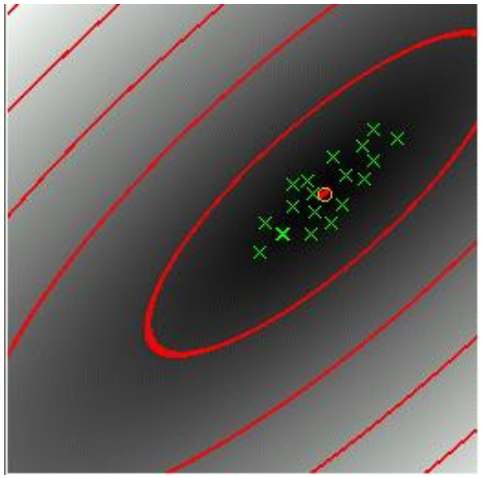
Cette distance ne tient compte ni des différences d'amplitude ni des éventuelles corrélations entre les différentes composantes des descripteurs.



La distance de Mahalanobis :

$$\delta_m(x, x')^2 = (x - x')^T C^{-1} (x - x')$$

Avec $C = \left(cov(x_i, x_j)\right)_{i,j}$ matrice de covariance de la base des descripteurs.



# APPARIEMENT DE PRIMITIVES : METRIQUES

Dans le cas de grande base de descripteurs (indexation), la matrice de covariance est calculée et mise à jour off-line. Si l'on diagonalise $C^{-1}$, on se ramène à un calcul de distance euclidienne sur des descripteurs normalisés :

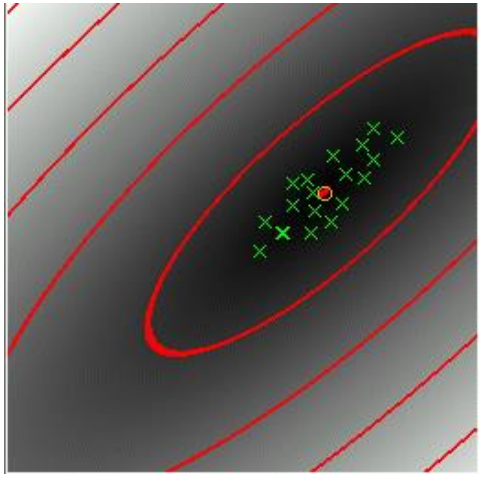
$$C^{-1} = P^T D P$$

$$\delta_m(x, x') = \sqrt{(x - x')^T C^{-1} (x - x')} = \underbrace{\left\| \sqrt{D} P x - \sqrt{D} P x' \right\|}$$

***distance ellipsoïdale***

A chaque mise à jour de la base d'index, on doit donc :

- Mettre à jour la matrice de covariance : $C$
- Calculer et diagonaliser : $C^{-1}$
- Normaliser tous les vecteurs : $x \longrightarrow \sqrt{D} P x$



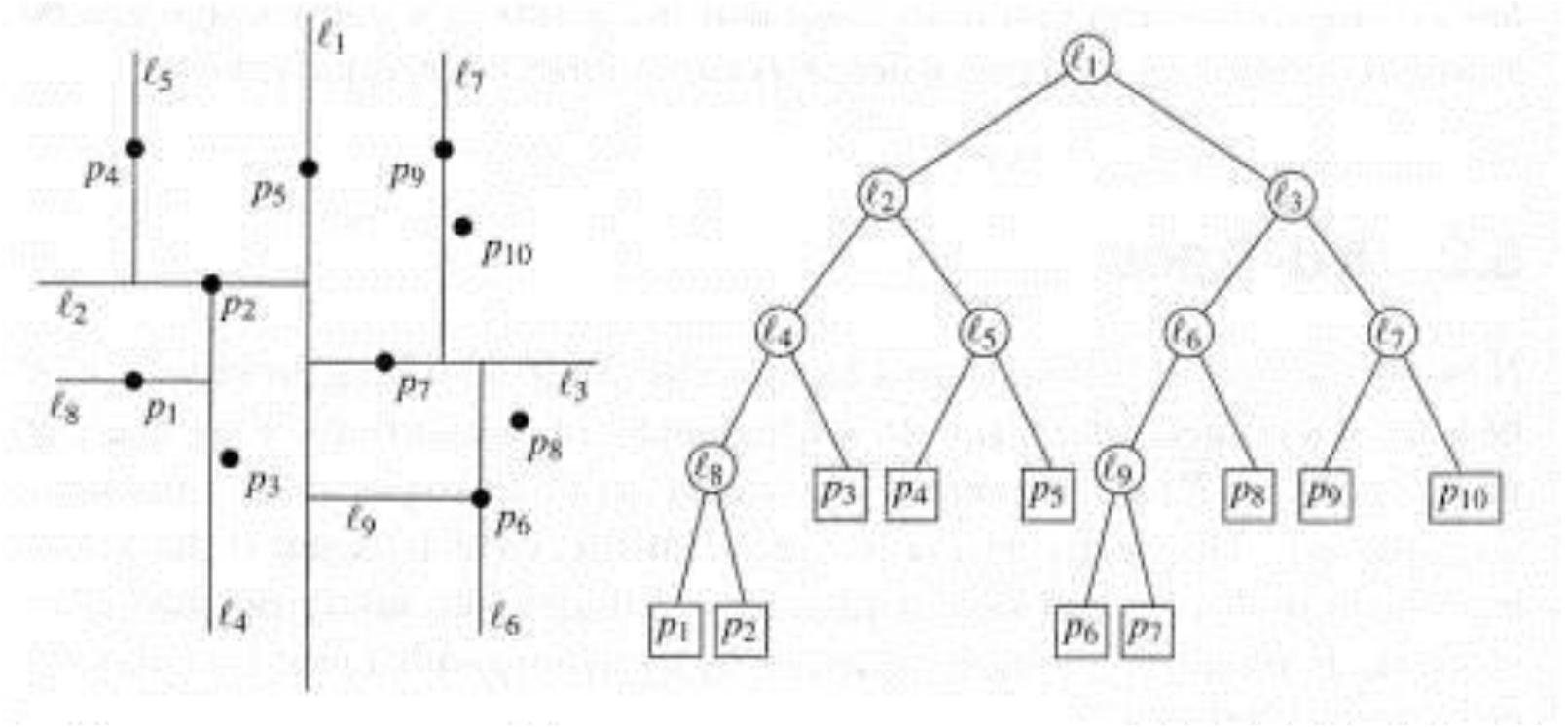
# BASE DE DESCRIPTEURS ET INDEXATION

Dans le cas d'une grosse base de descripteurs (indices), il est nécessaire de limiter la recherche à un certain voisinage du descripteur (index) inconnu. Ce problème est lié au stockage des vecteurs descripteurs dans la base d'index.

Découpage de la base de descripteurs en hypercubes

↕

Représentation de la base sous forme de Kd-tree



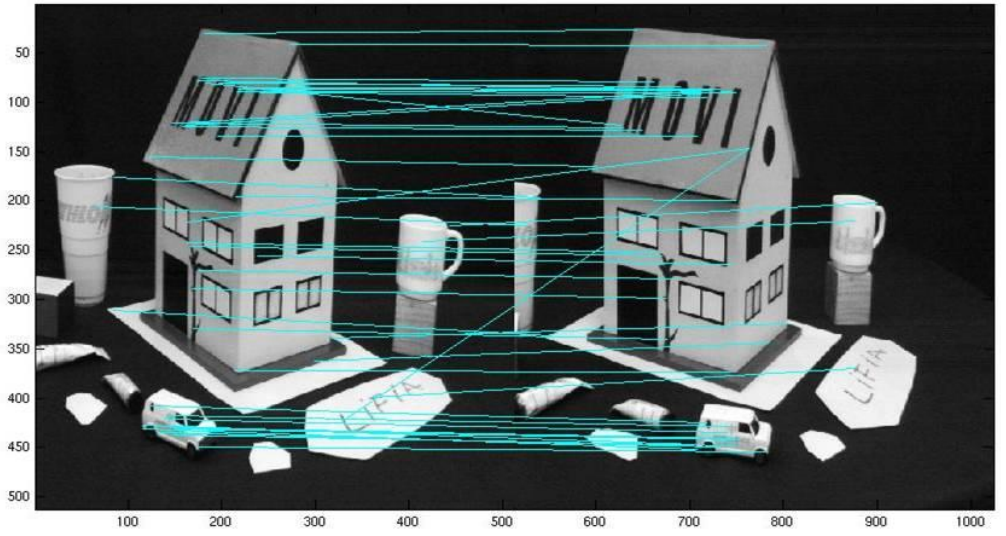
# DU LOCAL AU GLOBAL : CONSENSUS DES DESCRIPTEURS LOCAUX

Souvent l'utilisation des primitives visuelles aboutit à une décision prise à un niveau global sur l'image ou l'ensemble de points : étiquette de classe (reconnaissance, catégorisation), paramètres de déplacement (odométrie visuelle).

Comment réaliser cette décision collective à partir de l'ensemble des descripteurs ?

**Principe de vote** : chaque descripteur local est classifié et la classe globale est attribuée selon un critère majoritaire (ex : reconnaissance de pièces, catégories d'images…)

**Sélection par cohérence** : un sous-ensemble des appariements locaux est (itérativement) sélectionné pour établir une décision cohérente (ex : odométrie visuelle…)



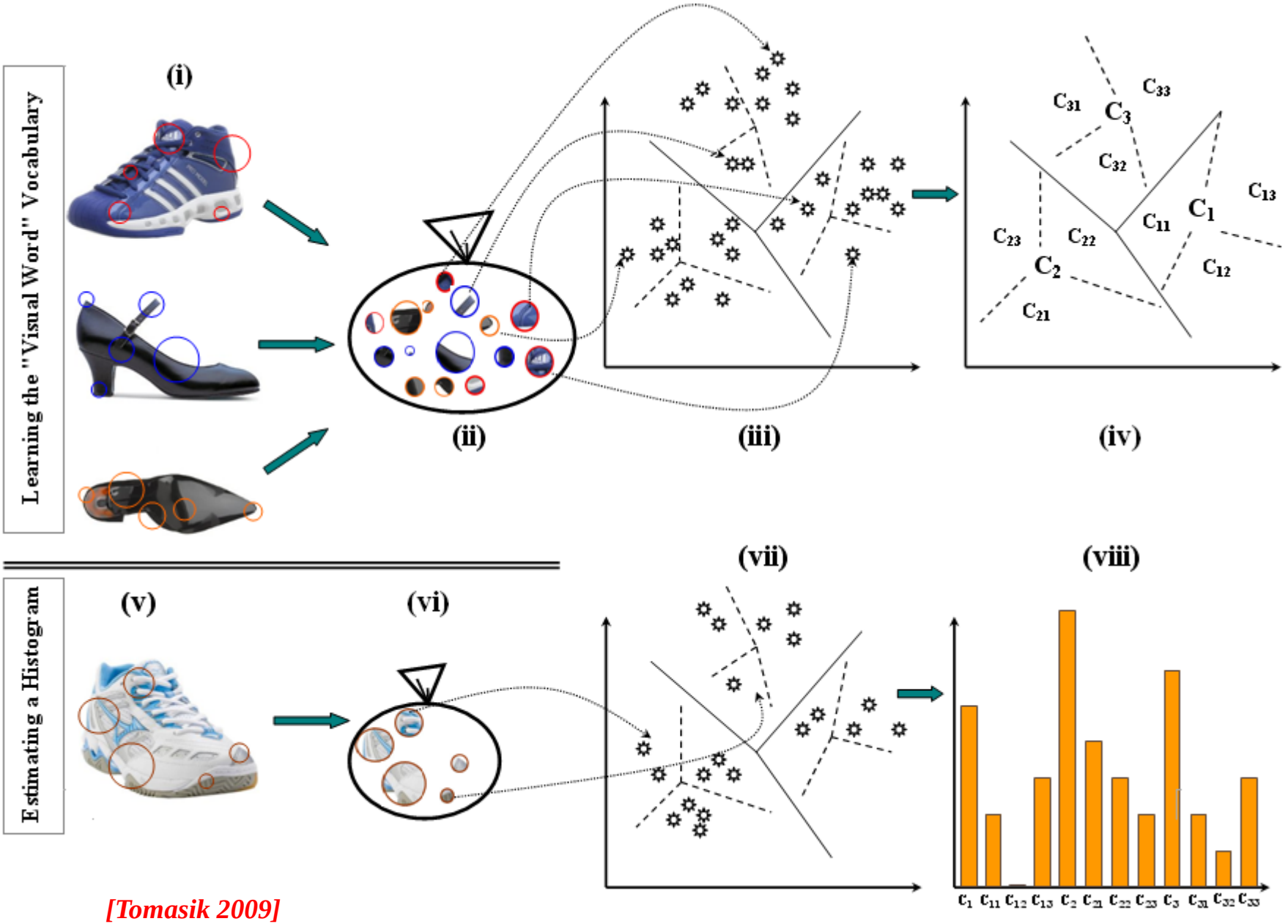
# DU LOCAL AU GLOBAL : SACS DE MOTS VISUELS

Une alternative classique aux techniques de consensus consiste à produire un descripteur global unique à partir de statistiques recueillies sur les descripteurs locaux :

- L'espace des descripteurs est réduit à un nombre limité d'étiquettes en utilisant un algorithme de quantification vectorielle (création d'un dictionnaire).
- Ce dictionnaire est utilisé pour coder les descripteurs locaux (classification locale, NN…)
- Un histogramme des « mots » est utilisé pour représenter globalement une image et la classifier.

*[Csurka 2004]*

# DU LOCAL AU GLOBAL : SACS DE MOTS VISUELS




*[Tomasik 2009]*

# APPARIEMENT GLOBAL : TECHNIQUES FREQUENTIELLES

Les techniques fréquentielles d'estimation du mouvement entre deux images sont fondées sur l'équivalence translation/déphasage de la transformée de Fourier :

Rappel : l'expression d'une image dans le domaine fréquentiel consiste à décomposer la fonction bidimensionnelle en sommes de sinusoïdes complexes :

$$I(x,y) = \frac{1}{wh}\sum_{u=0}^{w-1}\sum_{v=0}^{h-1} F(u,v)e^{2i\pi(ux+vy)/wh}$$

*Transformée de Fourier discrète inverse*

Les coefficients des différentes sinusoïdes sont calculés par la transformée de Fourier :

$$F(u,v) = \sum_{x=0}^{w-1}\sum_{y=0}^{h-1} I(x,y)e^{-2i\pi(ux+vy)/wh}$$

*Transformée de Fourier discrète directe*

Notation (module,phase) : $F(u,v) = \|F(u,v)\|e^{i\varphi_F(u,v)}$

La propriété de translation/déphasage dit que si $F$ est la transformée de Fourier de $I$ :

Alors la TF de $I$ translatée de $(-\delta x,-\delta y)$, est $G$, avec :

$$I(x,y) \xrightarrow{\text{TF}} F(u,v)$$

$$I(x+\delta x, y+\delta y) \xrightarrow{\text{TF}} G(u,v) = F(u,v)e^{2i\pi(u\delta x+v\delta y)/wh}$$

Soit : $\|G(u,v)\| = \|F(u,v)\|$ et : $\varphi_G(u,v) = \varphi_F(u,v) + 2\pi(u\delta x + v\delta y)/wh$

Le déphasage entre F et G vaut donc : $\Delta\phi(u,v) = 2\pi(u\delta x + v\delta y)/wh$

Il suffit donc en théorie de considérer ce déphasage pour 2 couples (u,v) pour calculer $(\delta x,\delta y)$, mais cette technique est sensible au bruit et aux changement d'illumination qui induisent des variations dans les basses fréquences.

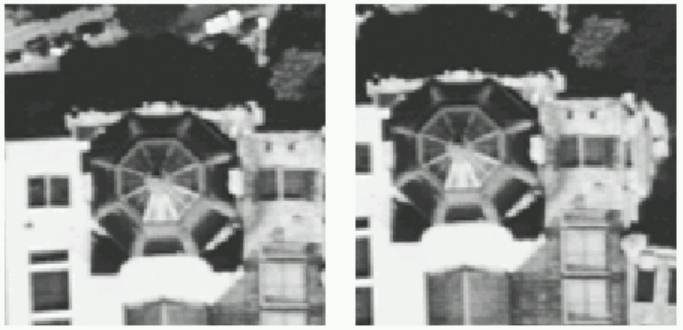
On utilise plutôt la technique de corrélation de phase.

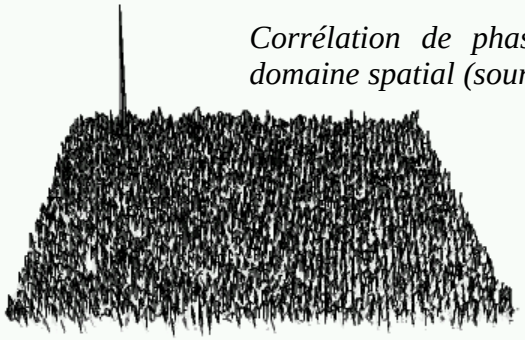
# CORRELATION DE PHASE

La technique de corrélation de phase exploite une conséquence directe de la propriété de translation/déphasage. Si $F$ est la TF de $I$ et $G$ la TF de $I$ translatée de $(-\delta x,-\delta y)$, alors le déphasage entre $F$ et $G$ est égal à leur spectre de puissance croisé normalisé (SPCN), i.e. :

$$\frac{F^*(u,v)G(u,v)}{\left\|F^*(u,v)G(u,v)\right\|} = e^{2i\pi(u\delta x+v\delta y)/wh}$$

La TF inverse du SPCN est donc égale à la fonction de Dirac du vecteur de translation : $\boldsymbol{\delta}_{(\delta x,\delta y)}(x,y)$




*Corrélation de phase dans le domaine spatial (source INRIA)*

La technique de corrélation de phase consiste donc à :

1. Calculer les TF de *I(x,y,t)* et *I(x,y,t+1)*, soit $F_1$ et $F_2$
2. Calculer $\chi$ le SPCN de *F1* et $F_2$
3. Calculer *D* la TF inverse de $\chi$
4. Rechercher le maximum de *D*

Avantages et inconvénients

+ Robuste car toutes les fréquences contribuent au calcul

+ Relativement rapide grâce au calcul de la FFT

- En pratique limité à un déplacement global sur toute l'image

# INVARIANTS DE FOURIER-MELLIN

L'utilisation de la transformée de *Fourier-Mellin* permet de calculer les paramètres d'une similitude (*rotation et homothétie*) comme un *vecteur de translation* de manière analogue au cas précédent, grâce à une représentation log-polaire de l'espace des fréquences $(u,v) \rightarrow (\theta, \log \rho)$ :

Soit $g$ l'image transformée de $f$ par une rotation d'angle $\alpha$, une homothétie de rapport $\sigma$, et une translation de vecteur $(x_0,y_0)$ :

$$g(x,y) = f\left(\sigma(\cos\alpha x + \sin\alpha y) - x_0, \sigma(-\sin\alpha x + \cos\alpha y) - y_0\right)$$

Les amplitudes des transformées de Fourier de $f$ et $g$ sont liées par la relation suivante :

$$\|G(u,v)\| = \frac{1}{\sigma^2}\left\|F\left(\tfrac{1}{\sigma}(u\cos\alpha + v\sin\alpha), \tfrac{1}{\sigma}(-u\sin\alpha + v\cos\alpha)\right)\right\|$$

donc l'amplitude :
- ne dépend pas de la translation $(x_0,y_0)$.
- subit une rotation d'angle $\alpha$.
- subit une modification d'échelle d'un facteur $1/\sigma$.

En passant les fréquences en coordonnées polaires :

$$F_p(\theta,\rho) = \|F(\rho\cos\theta, \rho\sin\theta)\|; 0 \le \theta \le 2\pi, 0 \le \rho < \infty$$
$$G_p(\theta,\rho) = \|G(\rho\cos\theta, \rho\sin\theta)\|; 0 \le \theta \le 2\pi, 0 \le \rho < \infty$$

on obtient :

$$G_p(\theta,\rho) = \frac{1}{\sigma^2} F_p\left(\theta - \alpha, \frac{\rho}{\sigma}\right)$$

Enfin, en passant la coordonnée radiale au logarithme :

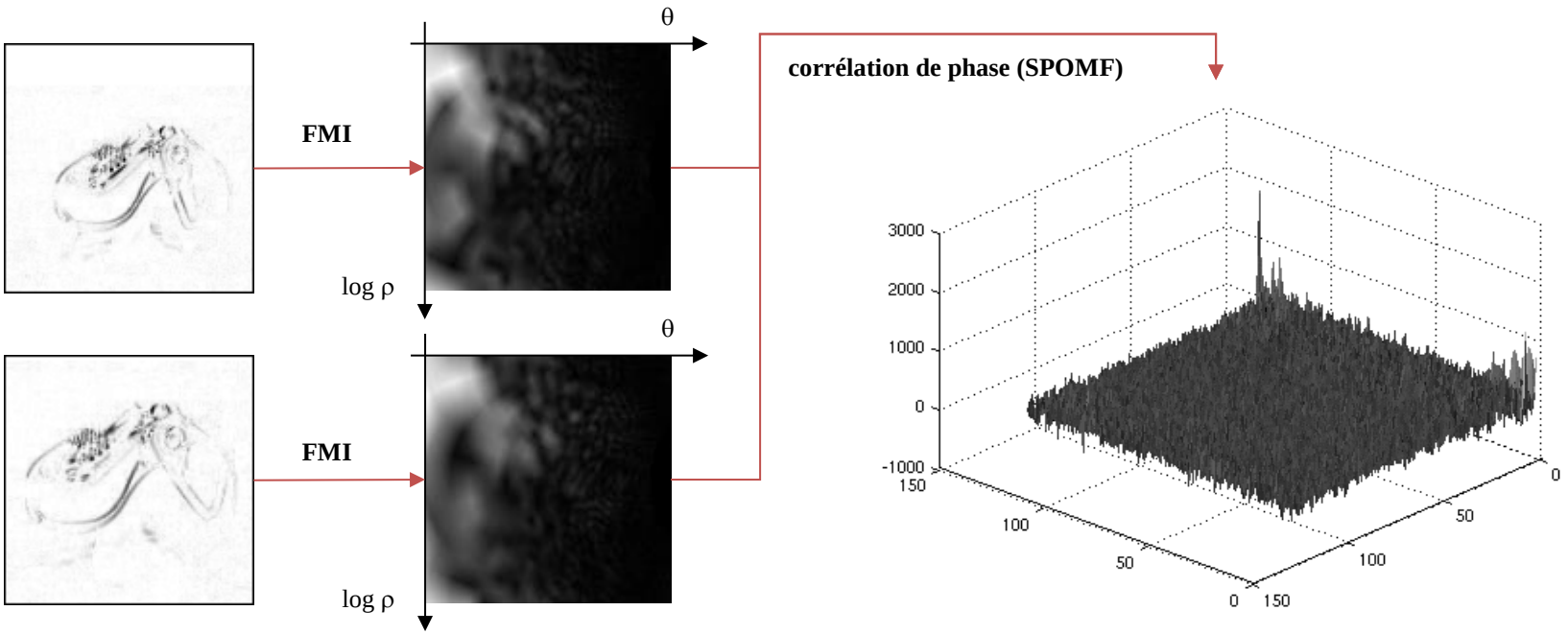
$$r = \log\rho \qquad F_{lp}(\theta,r) = F_p(\theta,\rho)$$
$$s = \log\sigma \qquad G_{lp}(\theta,r) = G_p(\theta,\rho)$$

on obtient :

$$\boxed{G_{lp}(\theta,r) = \frac{1}{\sigma^2} F_{lp}(\theta - \alpha, r - s)}$$

Donc une similitude dans l'espace image se traduit par une *translation* dans l'espace des *fréquences log-polaires*.

# INVARIANTS DE FOURIER-MELLIN : FMI-SPOMF




Un exemple d'utilisation de la transformée de Fourier-Mellin : calcul de la position de la tête des Robots Aibo dans l'image par corrélation de phase des invariants de Fourier-Mellin. (FMI-SPOMF : Fourier-Mellin Invariant Symmetric Phase Only Matched Filtering) : ***J.C. Baillie et M. Nottale*** 2004.

L'information de phase de l'image originale est perdue dans la FMI. Le FMI-SPOMF revient à chercher la meilleure (rotation, homothétie) qui mette en correspondance 2 spectres d'amplitude. *On ne retrouve donc pas les paramètres de translation entre les deux images, et de plus l'information de forme portée par la phase n'existe plus*.

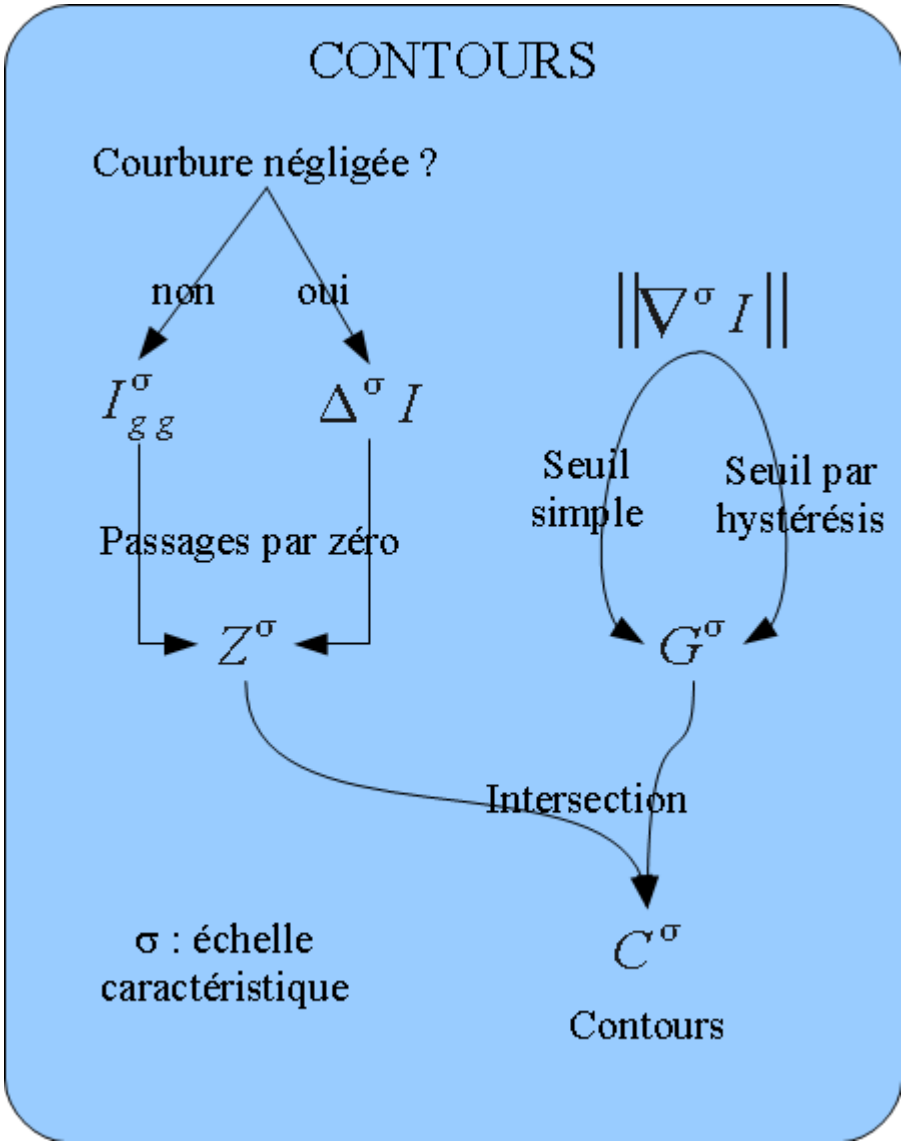
Pour compléter cette transformation, on peut appliquer une corrélation de phase classique sur le couple d'image de départ, après avoir appliqué sur l'une des images la transformation (rotation, homothétie) fournie par le FMI-SPOMF.

Notons enfin, que comme pour la corrélation de phase, cette méthode est utilisée en pratique pour estimer des transformations *globales*, car elle utilise la contribution de tout le spectre (ou au moins une large partie), ce qui implique une étendue spatiale importante des pixels utilisés pour l'estimation de chaque transformation.

# CONCLUSIONS : DERIVEES MULTI-ECHELLES ET CONTOURS

## DERIVEES MULTI-ECHELLES

- Dérivée estimée à une échelle donnée (variance de la gaussienne)
- Ordre 1, Gradient : Contraste, Direction…
- Ordre 2, Hessienne : Courbure, Contraste, Direction…
- Continuum du local (géométrie) vers le global (statistique).



# CONCLUSIONS : DETECTEURS ET DESCRIPTEURS

## DETECTEURS ET DESCRIPTEURS

**Détecteur** : réduction du support de calcul → répétable *et/vs* représentatif.

- Coins : Maxima de courbure, Harris, FAST…
- Blobs : Déterminant de la hessienne, SIFT, SURF…

**Descripteur** : représentation numérique → invariant *et/vs* discriminant.

- Invariants différentiels : couleur (intensité), contraste, laplacien,…
- Histogrammes d'indices contraste-invariants : direction, courbure,…

**Local** : géométrique → contour, courbure, coin, blob…

**Global** : statistique → histogramme, spectre d'amplitude/de phase…

Entre les deux : **analyse multi-échelles** → continuum…

# REFERENCES


- ***C. Harris & M. Stephens 1988*** ***« A combined corner and edge detector »*** *Alvey Vision Conference* ***pp 147-151***
- ***D.G. Lowe 2004*** ***« Distinctive Image Features from Scale-Invariant Keypoints »*** *International Journal of Computer Vision* ***60(2) pp 91-110***
- ***C. Schmid, R. Mohr & C. Bauckhage 2000*** ***« Evaluation of Interest Point Detectors »*** *Int. Jornal of Computer Vision* ***37(2) pp 151-172***
- ***C. Schmid & R. Mohr 1997*** ***« Local grayvalue invariants for image retrieval »*** *IEEE Transactions on Pattern Analysis and Machine Intelligence* ***19(5) pp 530-534***
- **E. Rosten & T. Drummond** "**Fusing points and lines for high performance tracking"** Int. Conf. on Computer Vision (ICCV 2005), 1508—1511, **2005.**
- **H. Bay, T. Tuytelaars & L. Van Gool** **"SURF: Speeded up robust features",** Computer Vision and Image Understanding, 110 (3), June, **2008,** 346-359

# REFERENCES


- ***N. Dalal & B. Triggs 2005*** ***« Histogram of oriented gradients for human detection », Int. Conf. Of Computer Vision and Pattern recognition (CVPR), 2005***
- **G. Csurka, C.R. Dance, L. Fan, J. Willamowski & C. Bray**, **"Visual categorization with bags of keypoints",** In Workshop on Statistical Learning in Computer Vision, ECCV, **2004.**
- ***B. Tomasik, P. Thiha & D. Turnbull 2009*** ***« Tagging products using image classification »,*** *SIGIR* ***2009.***
- ***H. Foroosh, J. Zerubia & M. Berthod 2002*** ***« Extension of phase correlation to subpixel registration »*** *IEEE Transactions on Image Processing* ***11(3) pp 188-200***
- ***Q. Chen, M. Defrise & F. Deconinck 1994*** ***« Symmetric Phase-Only Matched Filtering of Fourier-Mellin Transforms for Image Registration and Recognition »*** *IEEE Transactions on Pattern Analysis and Machine Intelligence* ***16(12) pp 1156-1168***