

MOUVEMENT

3

Recalage et Poursuite

PLAN DU CHAPITRE

- Recalage : méthodes fréquentielles
- Poursuite 1 : Harris et Invariants de Hilbert
- Poursuite 2 : SIFT

Introduction

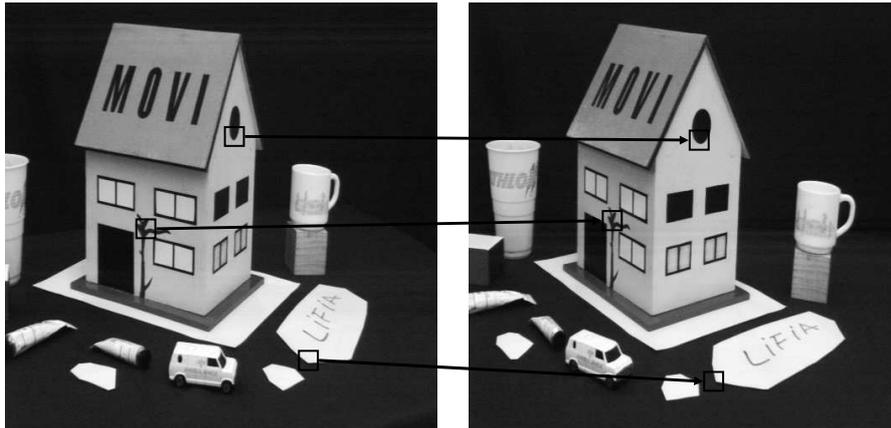


Lorsque l'amplitude du déplacement est grande, l'estimation directe par les méthodes d'appariement ou les méthodes différentielles est vouée à l'échec, pour 2 raisons majeures :

(1) Le déplacement apparent peut être supérieur aux périodes spatiales présentes dans les images (aliasing temporel, voir exemple ci-contre)

Introduction

(2) La scène est soumise à des déformations géométriques complexes qui invalident l'hypothèse de translation locale (voir exemple ci-dessous)



Cours ETASM 2006

Antoine MANZANERA - ENSTALEI

3

Introduction

Par conséquent :

- Il est primordial de disposer de descripteurs qui soit le plus *invariants* possible aux *transformations géométriques* : rotation, homothétie, transformation affine. Cela passe forcément par un calcul multi-échelle.
- Les descripteurs fournis auront un caractère beaucoup *moins local*, voire global, et on ne disposera plus d'un champ dense comme dans les méthodes d'estimation, mais au mieux d'un *champ épars*.

Ce qu'on verra dans ce chapitre :

TRANSFORMATION GLOBALE :

- Un seul paramètre de déplacement pour toute l'image (translation, rotation, homothétie) : *Méthodes fréquentielles*.

MOUVEMENT QUELCONQUE :

- Quels sont les points utilisés pour le calcul ? *Points de Harris, Maxima locaux dans l'espace d'échelle,...*
- Quels descripteurs utiliser pour l'appariement ? *Invariants différentiels, histogrammes d'orientation,...*

Applications : Reconnaissance d'objets, Indexation d'images, localisation de robots,...

Cours ETASM 2006

Antoine MANZANERA - ENSTALEI

4

Techniques fréquentielles

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 sinusoides complexes :

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

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

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

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

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 :

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

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

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 changements d'illumination qui induisent des variations dans les basses fréquences.

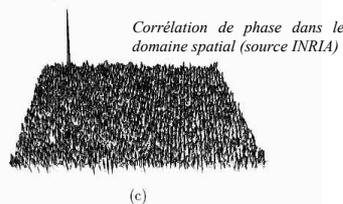
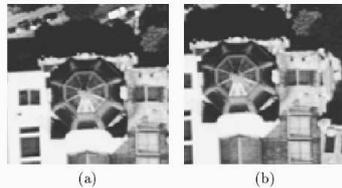
On utilise plutôt la technique de corrélation de phase.

Corrélation 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. :

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

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



La technique de corrélation de phase consiste donc à :

1. Calculer les TF de $I(x, y, t)$ et $I(x, y, t+I)$, soit F_1 et F_2
2. Calculer χ le SPCN de F_1 et F_2
3. Calculer D la TF inverse de χ
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

Invariant 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 α , une homothétie de rapport σ , et une translation de vecteur (x_0, y_0) :

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

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

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

donc l'amplitude :

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

En passant les fréquences en coordonnées polaires :

$$F_p(\theta, \rho) = \|F(\rho \cos \theta, \rho \sin \theta)\|, 0 \leq \theta \leq 2\pi, 0 \leq \rho < \infty$$

$$G_p(\theta, \rho) = \|G(\rho \cos \theta, \rho \sin \theta)\|, 0 \leq \theta \leq 2\pi, 0 \leq \rho < \infty$$

Enfin, en passant la coordonnée radiale au logarithme :

$$r = \log \rho$$

$$s = \log \sigma$$

$$F_p(\theta, r) = F_p(\theta, \rho)$$

$$G_p(\theta, r) = G_p(\theta, \rho)$$

on obtient :

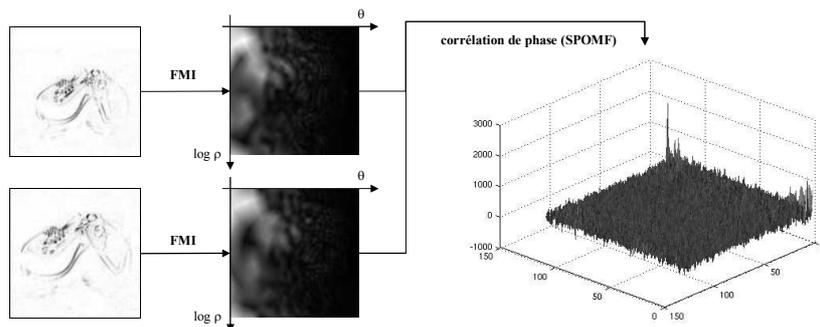
$$G_p(\theta, r) = \frac{1}{\sigma^2} F_p\left(\theta - \alpha, \frac{\rho}{\sigma}\right)$$

on obtient :

$$G_p(\theta, r) = \frac{1}{\sigma^2} F_p(\theta - \alpha, r - s)$$

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

Invariant de Fourier-Mellin



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. Baillic et M. Nottale* 2004.



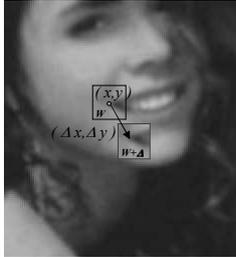
L'information de phase de l'image originale est perdue dans la FMI. Le FMI-SPOMF revient à chercher la meilleure (rotation, homothétie) qui mette en correspondance 2 spectres d'amplitude. *On ne retrouve donc pas les paramètres de translation entre les deux images, et de plus l'information de forme portée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

Détection de points anguleux

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* :

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

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

Or, en utilisant une approximation du premier ordre :

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

Et donc :

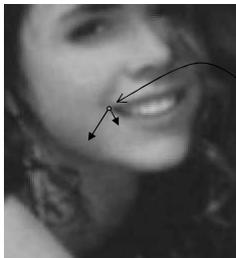
$$\chi(x, y) = \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial x}(x_k, y_k) \cdot \frac{\partial I}{\partial y}(x_k, y_k) \right) \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}^2 = (\Delta x \quad \Delta y) \underbrace{\begin{pmatrix} \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial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

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

Détection de points anguleux

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

La matrice d'autocorrélation Ξ 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 $(\Delta x, \Delta y) \cdot \Xi(x, y) \cdot (\Delta x, \Delta y)^t$ 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 Ξ décrivant des variations locales importantes de l'image.

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

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

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 Θ qui sont au delà d'un certain seuil (typiquement 1% de la valeur max de Θ).

[Harris 88]

Utilisation des invariants différentiels

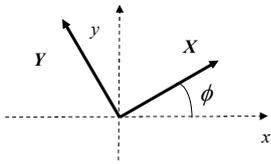
Objectif : représenter les points d'intérêt par des indices qui soient invariants par rotation et par changement d'échelle.

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

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



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

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

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

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

Utilisation des invariants différentiels

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

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

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

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

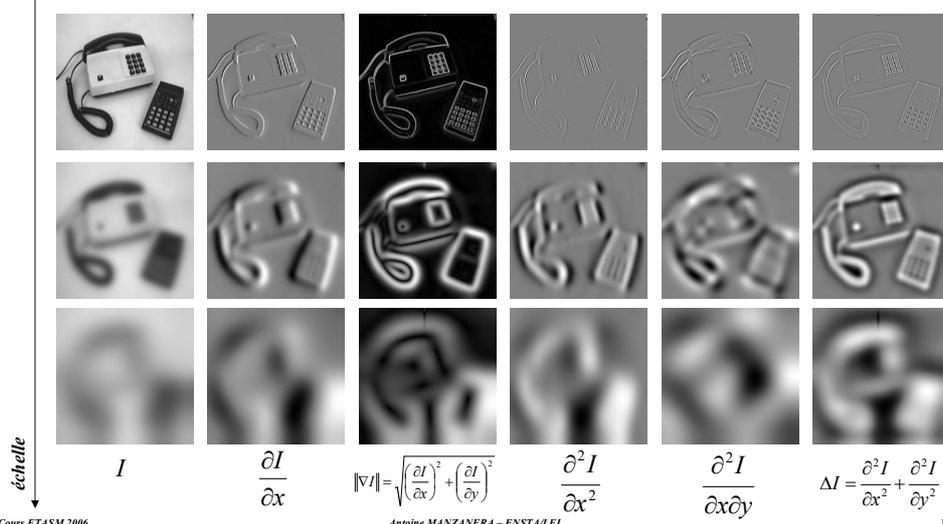
$$\begin{aligned} \Psi_2 &= L_i L_j L_j = L_x L_{xx} L_x + 2L_x L_{xy} L_y + L_y L_{yy} L_y \\ \Psi_7 &= -\varepsilon_{ij} L_{jkl} L_i L_k L_l = L_{xxy} (-L_x L_x L_x + 2L_x L_y L_y) \\ &+ L_{xyy} (-2L_x L_x L_y + L_y L_y L_y) - L_{yyy} L_x L_y L_y + L_{xxx} L_x L_x L_y \end{aligned}$$

NB : invariance par rotation du noyau gaussien !

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

Espace d'échelle gaussien et dérivées

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{\pi}} \exp\left(-\frac{x^2}{\sigma^2}\right) \quad g_\sigma \quad \frac{\partial g_\sigma}{\partial x} \quad \frac{\partial^2 g_\sigma}{\partial x^2} \quad \partial^n (I * g_\sigma) = I * \partial^n g_\sigma$$



Mise en œuvre du détecteur de Harris

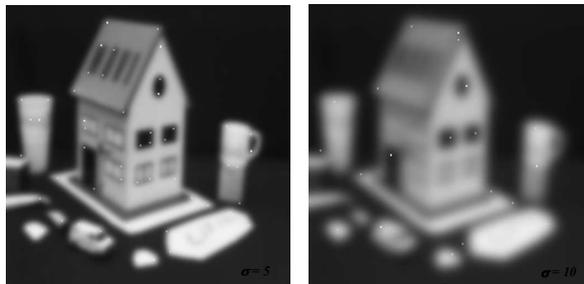
1. On calcule les dérivées premières à partir des dérivées de gaussienne (écart-type σ_1)
2. On calcule les termes de la matrice d'autocorrélation Ξ en calculant une moyenne locale des dérivées sous la forme d'une gaussienne (écart-type σ_2 , typiquement $\sigma_2 = 2 \sigma_1$)
3. On calcule la fonction d'intérêt : $\Theta = \det(\Xi) - \alpha \text{trace}(\Xi)$ (typiquement $\alpha = 0,06$).
4. On calcule les maxima locaux de Θ supérieurs à un certain seuil (typiquement 1% de Θ_{\max}).



Détecteur de Harris multi-échelles



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

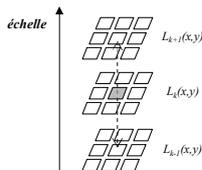
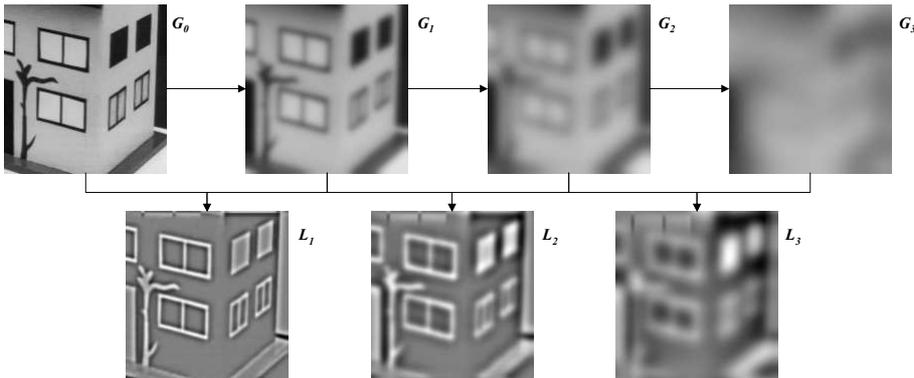


Cours ETASM 2006

Antoine MANZANERA - ENSTALEI

15

Détecteur SIFT : extrema dans l'espace d'échelle



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).

Cours ETASM 2006

Antoine MANZANERA - ENSTALEI

16

Points d'intérêt SIFT

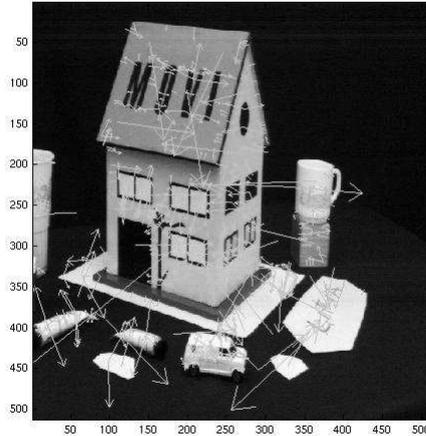


Image 1 : 589 points détectés.

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

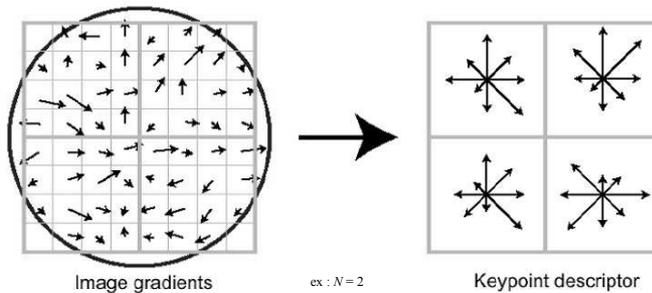
(où σ est l'échelle sélectionnée)

Ci-contre, point d'intérêt SIFT : la direction de la flèche représente la direction θ et sa longueur l'échelle σ associée.

Descripteur SIFT : histogramme 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 4×4 .
- On calcule le gradient $(G_x(a, b, \sigma), G_y(a, b, \sigma))$ pour les $4 \times 4 \times N^2$ points (a, b) .
- Pour chaque carré 4×4 , 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.



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

Appariement par 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.

Conclusion Chapitre 3

- ❑ RECALAGE : Mouvement global
 - ❑ Corrélation de phase : *translation*
 - ❑ Invariants de Fourier-Mellin : *rotation / homothétie*
- ❑ POURSUITE
 - ❑ Points d'intérêt
 - ❑ Harris : *points anguleux*
 - ❑ SIFT : *extrema espace d'échelle*
 - ❑ Descripteurs
 - ❑ Invariants différentiels
 - ❑ Histogrammes d'orientation

Bibliographie du Chapitre 3

- **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
- **C. Harris & M. Stephens 1988** « *A combined corner and edge detector* » *Alvey Vision Conference* pp 147-151
- **C. Schmid & R. Mohr 1997** « *Local grayvalue invariants for image retrieval* » *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19(5)** pp 530-534
- **D.G. Lowe 2004** « *Distinctive Image Features from Scale-Invariant Keypoints* » *International Journal of Computer Vision* **60-2** pp 91-110