Poursuite de cible.

1• Qu'est-ce que la poursuite de cible ?

Lorsqu'un objet bouge devant une caméra ou lorsque la caméra bouge, il se produit une variation de l'illumination du capteur d'image (rétine) modifiant la distribution des niveaux de gris (voir Figure 1).

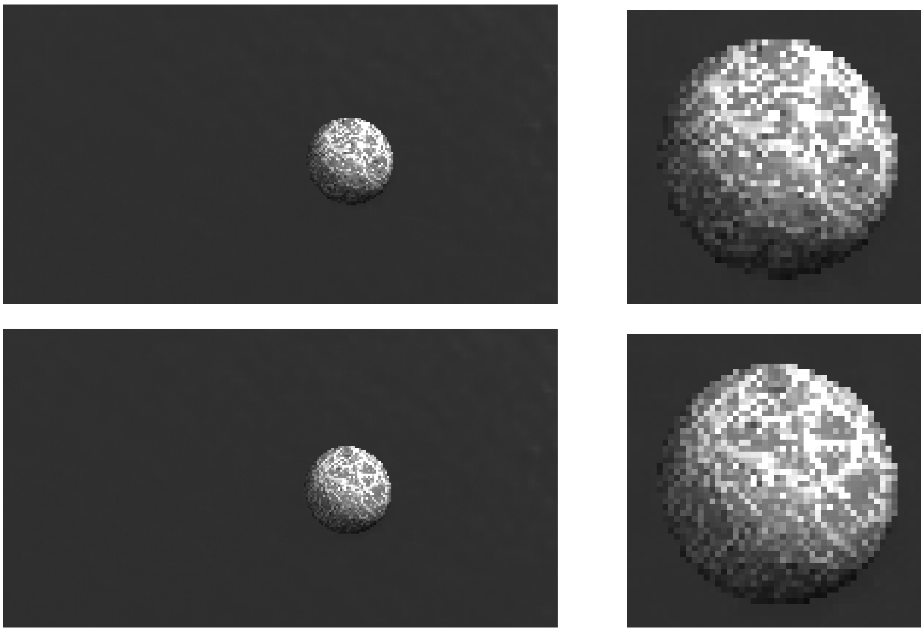


Figure 1 : deux images consécutives de la séquence "Boule de marbre" et détails.

Cette modification des niveaux de gris porte le nom de "mouvement apparent". Le mouvement apparent est l'effet que produit un mouvement réel sur la distribution de l'illumination projetée.

La poursuite de cible est une technique permettant de trouver, dans une séquence d'images, la position d'une zone prédéfinie correspondant à un motif particulier dont on souhaite connaître la position.

Les utilisations de la poursuite de cible sont multiples :

- asservissement d'un robot sur une cible,
- caractérisation du fonctionnement d'un organe en imagerie médicale,
- vidéo surveillance automatique,
- contrôle du fonctionnement d'une canalisation,
- suivi de planète en astronomie,
- ...

Pour réaliser une poursuite de cible, il faut être capable de retrouver le mouvement projeté de la cible à partir des variations des niveaux de gris de la séquence d'images. Plusieurs méthodes peuvent être utilisées pour ce faire qui vous sont détaillées dans le paragraphe suivant.

2• MÉTHODES DE MISE EN CORRESPONDANCE DE MOTIF.

2.1• Principe.

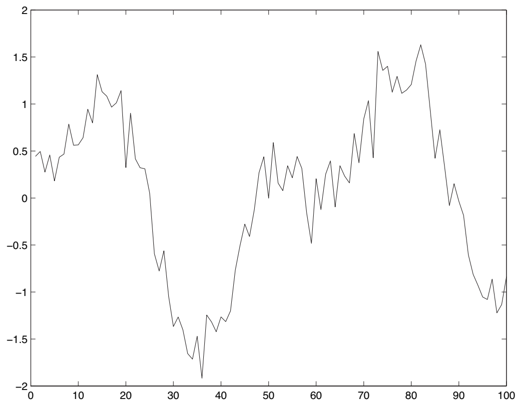
Le principe de la mise en correspondance de motif repose sur une mesure de similarité entre deux images. Ces mesures de similarités sont basées soit sur des distances statistiques, soit sur des mesures de corrélation.

2.2• Corrélation.

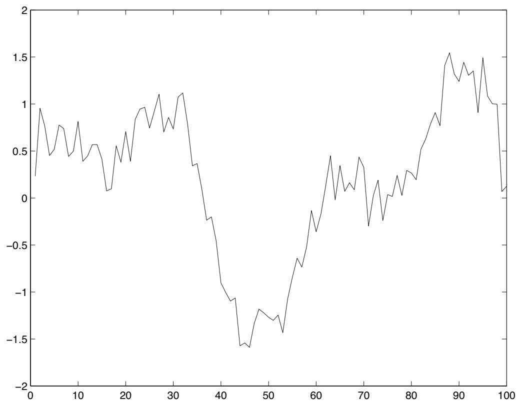
La corrélation de deux signaux échantillonnés x_k et y_k est donnée par :

$$R_{xy}(\tau) = \frac{1}{N}\sum_{k=1}^{N} x_k y_{k+\tau}.$$

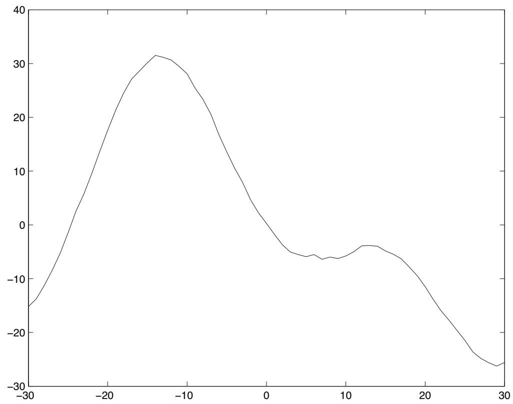
Si x_k et y_k sont deux réalisations (avec des variations aléatoires) d'un même processus, alors $R_{xy}(\tau)$ passe par un maximum pour une valeur donnée T de τ.



signal x_k



signal y_k



Corrélation entre x et y

Sur les figures ci-dessus, on voit que la corrélation entre x et y passe par un maximum lorsque $\tau = -12$ qui est le décalage entre x_k et y_k.

Attention ! Mathématiquement, on écrit les corrélations sans se préoccuper des indices manipulés. En général on ne dispose que d'échantillons bornés. Si N est le nombre d'échantillons x_k et y_k, on ne pourra calculer R_{xy} qu'avec un nombre réduit de valeurs. En supposant que τ soit recherché dans l'intervalle [-d, d], alors la formule du calcul de R_{xy}

doit être modifiée de la façon suivante :

$$R_{xy}(\tau) = \frac{1}{N-2d} \sum_{k=d+1}^{N-d} x_k y_{k+\tau}$$

Pour vous persuader de ça, essayer (à la main) de calculer R_{xy} dans l'intervalle [-2,2] pour les deux séries suivantes :

$x_{(k)}$ = 1, 3, 5, 1, 2, 2, 2, 4, 2, 2

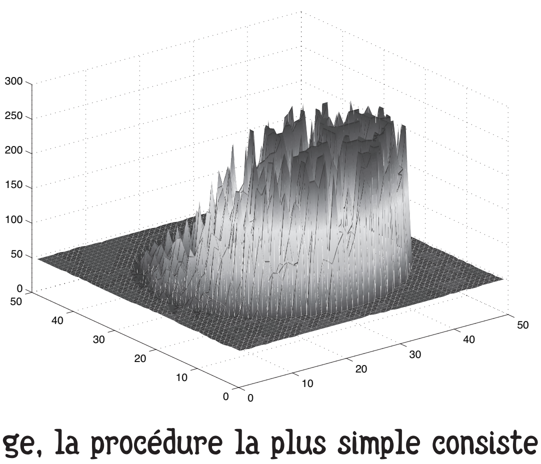
$y_{(k)}$ = 4, 5, 6, 5, 2, 3, 6, 6, 3, 5

Lorsque les deux signaux x_k et y_k n'ont pas la même moyenne, le décalage des moyennes peut induire de faux maxima dans la fonction de corrélation. C'est pourquoi on préfère utiliser la corrélation centrée :

$r_{xy}(\tau) = \frac{1}{N} \sum_{k=1}^{N} (x_k - \bar{x})(y_{k+\tau} - \bar{y})$ où $\bar{x}$ et $\bar{y}$ sont respectivement les moyenne des processus x_k et y_k : $\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k$, $\bar{y} = \frac{1}{N} \sum_{k=1}^{N} y_k$.

Enfin, l'auto-corrélation d'un signal avec lui-même (avec un décalage τ de 0) étant toujours supérieure à la corrélation de ce signal avec n'importe quel autre signal, cette valeur peut être utilisée en référence pour produire une valeur normalisée, appelée coefficient de corrélation. Le coefficient de corrélation vaut 1 si les deux signaux x_k et $y_{k+\tau}$ se superposent parfaitement et est inférieur à 1 sinon. Il s'écrit :

$$\rho_{xy}(\tau) = \frac{\sum_{k=1}^{N} (x_k - \bar{x})(y_{k+\tau} - \bar{y})}{\sqrt{\sum_{k=1}^{N} (x_k - \bar{x})^2 \sum_{k=1}^{N} (y_k - \bar{y})^2}}$$



L'utilisation de la corrélation pour la poursuite de motif consiste à supposer que les deux images à mettre en correspondance (le motif recherché et l'image courante) sont deux signaux bi-dimensionnels. Pour retrouver la position du motif sur l'image courante, il faut retrouver le décalage (u,v) donnant la meilleure corrélation entre le signal *Motif* et le signal *Image courante*. Pratiquement, le motif à rechercher étant toujours d'une dimension inférieure à celle de l'image, la procédure la plus simple consiste à extraire de l'image courante des imagettes de même taille que le motif sélectionné et de calculer la corrélation.

En traitement d'image la mesure de corrélation 2D est généralement notée CC.
Son écriture est la suivante :

$$CC(u, v) = \frac{1}{N}\sum_{i}\sum_{j} I_1(i, j) I_2(i+u, j+v) = \frac{1}{N}\sum_{i,j} I_1(i, j) I_2(i+u, j+v)$$

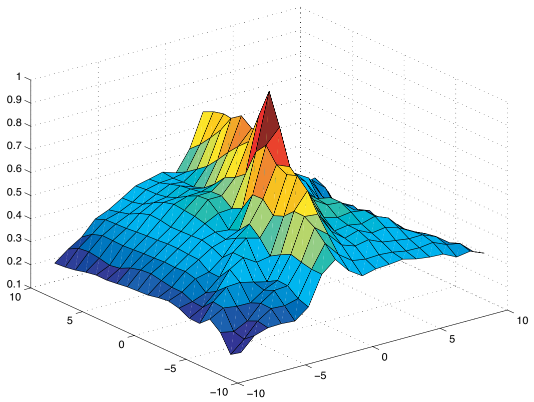
où N est le nombre d'éléments de la somme. On note généralement la double somme sous la forme d'une simple somme.
La version centrée est notée ZCC (Z comme zéro) :

$$ZCC(u, v) = \frac{1}{N}\sum_{i,j}(I_1(i, j) - \bar{I}_1)(I_2(i+u, j+v) - \bar{I}_2)$$

Enfin le coefficient de corrélation se note NZCC (N pour normalisé)

$$ZNCC(u, v) = \frac{\sum_{i,j}(I_1(i, j) - \bar{I}_1)(I_2(i+u, j+v) - \bar{I}_2)}{\sqrt{\sum_{i,j}(I_1(i, j) - \bar{I}_1)^2 \sum_{i,j}(I_2(i, j) - \bar{I}_2)^2}}$$

Les valeurs de u et v qui maximisent cette fonction de corrélation doivent être retenues comme mesure du décalage entre le motif et l'image.



2.3• Distance absolue.

Il est aussi possible de mesurer une distance absolue entre deux images. Cette distance porte le nom de SAD (somme absolue des distances). Sa formulation est :

$$SAD(u, v) = \frac{1}{N}\sum_{i,j}\left|I_1(i, j) - I_2(i+u, j+v)\right|$$

La version centrée est :

$$ZSAD(u, v) = \frac{1}{N}\sum_{i,j}\left|(I_1(i, j) - \bar{I}_1) - (I_2(i+u, j+v) - \bar{I}_2)\right|$$

Enfin en normalisé :

$$ZNSAD(u, v) = \frac{\sum_{i,j}\left|(I_1(i, j) - \bar{I}_1) - (I_2(i+u, j+v) - \bar{I}_2)\right|}{\sqrt{\sum_{i,j}\left|I_1(i, j) - \bar{I}_1\right| \sum_{i,j}\left|I_2(i, j) - \bar{I}_2\right|}}$$

2.4• Distance Euclidienne.

C'est une autre distance statistique appelée somme des écarts quadratiques (summ of square distances) :

$$SSD(u, v) = \frac{1}{N}\sum_{i,j}(I_1(i,j) - I_2(i+u, j+v))^2$$

Somme des écarts quadratiques centrés :

$$ZSSD(u, v) = \frac{1}{N}\sum_{i,j}((I_1(i,j) - \bar{I}_1) - (I_2(i+u, j+v) - \bar{I}_2))^2$$

Somme des écarts quadratiques centrés normalisée :

$$ZNSSD(u, v) = \frac{\sum_{i,j}((I_1(i,j) - \bar{I}_1) - (I_2(i+u, j+v) - \bar{I}_2))^2}{\sqrt{\sum_{i,j}(I_1(i,j) - \bar{I}_1)^2 \sum_{i,j}(I_2(i+u, j+v) - \bar{I}_2)^2}}$$

2.5• Distance robuste (méthodes par rang).

Les méthodes par distance statistiques de type L1 (valeur absolue) ou L2 (distance euclidienne) ont parfois des problèmes de robustesse. On lui préfère alors des méthodes agissant par tri connues pour être plus robustes.

• **Rank**

La méthode consiste à remplacer le niveau de gris d'un pixel par son rang puis à utiliser des distances de rang comme la distance de Spearman ou celle de Kendall.

Spearman : $\rho = 1 - \frac{6}{n^3 - n}\sum_i (R_i - Q_i)^2$ où R_i est le rang dans la première image et Q_i le rang dans la seconde image.

Kendall : on somme le nombre de concordance, c'est à dire le nombre de fois ou on a à la fois i<j et $R_i < R_j$. Soit C cette somme de n.(n-1) concordances : $\rho = \left|2C - \frac{n(n-1)}{2}\right|$

Plus ρ est grand, plus les imagettes sont corellées.

Exemple :

10	16	30	18
12	19	20	22
15	11	21	26
27	13	28	25

les rangs de cette image sont :

1	6	16	7
3	8	9	11
5	2	10	13
14	4	15	12

ce qui développé en ligne donne :

1	6	16	7	3	8	9	11	5	2	10	13	14	4	15	12

une autre image peut donner :

1	8	15	6	4	9	7	16	5	3	11	13	12	2	10	14

$\sum_i (R_i - Q_i)^2 = 76$ et $\rho = 0{,}888$ pour Spearman.

C = 42 pour Kendall et $\rho = 36$

• **Census**

Il s'agit ici de créer un motif binaire par comparaison avec le pixel central du motif :

10	16	30
12	19	20
15	11	21

La comparaison donne:

0	0	1
0	0	1
0	0	1

ce qui développé en ligne donne :

0	0	1	0	0	1	0	0	1

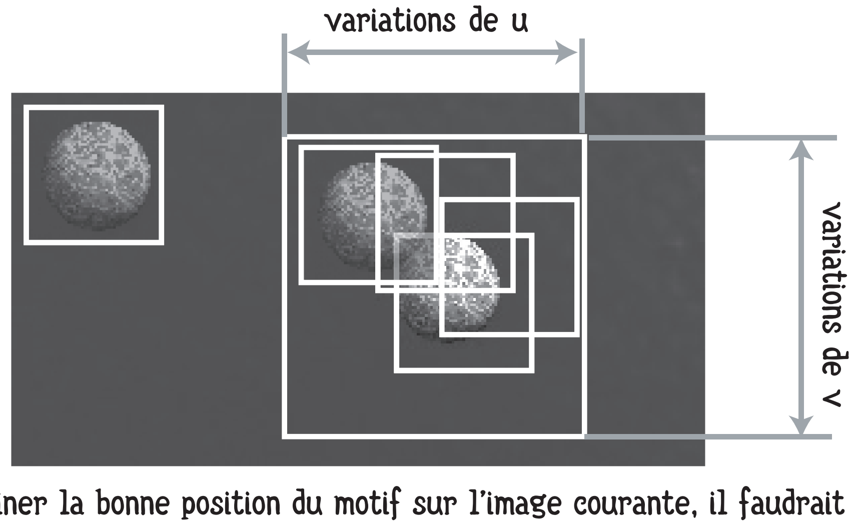
une autre imagette aurais pu donner :

0	1	1	1	0	0	0	0	1

La distance est alors donnée par une somme de ou-exclusifs sur ces valeurs binaires.

Dans le cas présent la distance est de $\frac{3}{9}$ =0,33.

2.6• D'un point de vue pratique.



Pour déterminer la bonne position du motif sur l'image courante, il faudrait normalement tenter de positionner le coin en haut à gauche du cadre du motif en chaque point de l'ima-

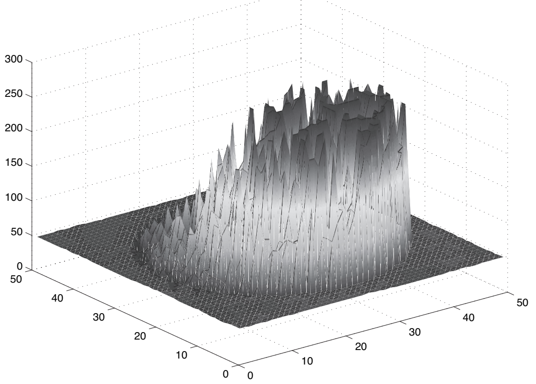
ge. C'est cette solution qui est retenue dans l'exemple qui vous est fourni dans le fichier *Poursuite.m*. Cette solution est particulièrement lente (comme vous pourrez vous en rendre compte).
En pratique, si on sait que le mouvement ne peut excéder un certaine valeur de D pixels, il suffit de positionner le cadre dans toutes les positions définies par un cadre de largeur 2D autour de la position d'origine de l'image.
Donc la procédure est la suivante :
pour chaque valeur de u et de v dans l'intervalle des mouvements possibles
• on extrait de l'image courante une imagette de la même taille que le motif
• on calcule l'indice de similitude entre cette imagette et le motif (corrélation, SAD, ...)
Une fois que toutes les valeurs de similitude sont calculée, on retient celle dont la valeur est maximale (dans le cas de la corrélation) ou minimale (dans le cas des distances).

3• Déplacement de la fenêtre de référence.

Le temps de calcul d'une corrélation ou d'une distance statistique entre deux motifs (le motif de référence et l'imagette extraite) est proportionnel à la taille du motif de référence et à la taille de la fenêtre de recherche. Si le motif à poursuivre se déplace dans toute l'image, la fenêtre de recherche doit couvrir (a priori) toute l'image. On se retrouve dans le mauvais cas précédent.
Cependant, entre deux images successives, on peut supposer que le motif ne s'est pas déplacé d'un bout à l'autre de l'image. Dans ce cas, on peut sélectionner, pour la fenêtre de recherche, une zone autour de la position du motif dans l'image précédente. On obtient alors un véritable algorithme de poursuite de motif.

4• Flot Optique.

La méthode du flot optique consiste à supposer que le motif est une fonction discrète à deux dimensions produite par la discrétisation (par le capteur d'images) de l'illumination locale.



Soit I(x,y,t) la répartition de l'illumination sur la fenêtre courante. Si le motif est correctement recalé, I(x,y,t) est aussi la fonction d'illumination du motif.
A l'instant suivant le motif s'est déplacé de (dx, dy). Soit dt le laps de temps écoulé entre deux acquisitions d'images.
Si le motif a toujours la même répartition d'illumination, alors $I(x+dx,y+dy,t+dt)=I(x,y,t)$.
La méthode du flot optique consiste à faire une expansion au premier ordre de cette équation :

$$I(x+dx,\ y+dy,\ t+dt) = I(x,y,t) + \frac{\partial}{\partial x}I(x,y,t)dx + \frac{\partial}{\partial y}I(x,y,t)dy + \frac{\partial}{\partial t}I(x,y,t)dt + O^2$$

ou O^2 regroupe les termes d'ordre supérieurs.

Et comme la conservation de l'illumination impose $I(x+dx,y+dy,t+dt)=I(x,y,t)$, il vient :

$$\frac{\partial}{\partial x}I(x,y,t)dx + \frac{\partial}{\partial y}I(x,y,t)dy + \frac{\partial}{\partial t}I(x,y,t)dt \approx 0$$

ou encore :

$$\frac{\partial}{\partial x}I(x,y,t)\frac{dx}{dt} + \frac{\partial}{\partial y}I(x,y,t)\frac{dy}{dt} + \frac{\partial}{\partial t}I(x,y,t) \approx 0$$

$\frac{\partial}{\partial t}I(x,y,t)$, c'est la variation de l'illumination au point (x,y) entre les deux instant t et t+dt. C'est donc (approximativement) $\frac{\partial}{\partial t}I(x,y,t) = I_t(x,y,t) = I(x,y,t+dt) - I(x,y,t)$.

Quant aux quantités $I_x(x,y) = \frac{\partial}{\partial x}I(x,y,t)$ et $I_y(x,y) = \frac{\partial}{\partial y}I(x,y,t)$, ce sont des valeurs qui peuvent être calculées a priori sur l'image grâce à un opérateur que l'on nomme "extracteur de gradient". L'extracteur de gradient permet d'obtenir, à partir d'une image $I(x,y,t)$, les valeurs estimées des deux images $I_x(x,y)$ et $I_y(x,y)$ en chaque point du motif.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
16	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

Si on suppose que tous les pixels d'un motif sont numérotés comme ci-contre (on suppose ici que le motif est (6,6) et on a donc 36 pixels de coordonnées $(x_i, y_i)_{i \text{ de } 1 \text{ à } 36}$). Si on suppose que tous les points du motif subissent le même déplacement (u,v), alors on dispose, pour identifier (u,v) de 36 équations de flot optique pour deux inconnues :

$$\begin{bmatrix} I_x(x_1, y_1) & I_y(x_1, y_1) \\ \dots & \dots \\ I_x(x_{36}, y_{36}) & I_y(x_{36}, y_{36}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} I_t(x_1, y_1, t) \\ \dots \\ I_t(x_{36}, y_{36}, t) \end{bmatrix}$$

Cette équation est de la forme : $A\begin{bmatrix} u \\ v \end{bmatrix} = B$. Elle peut être résolue par moindres carrés en calculant $A^\dagger$ la pseudo-inverse de A. Comme A ne varie pas, on peut précalculer $A^\dagger$.

La méthode de calcul est alors la suivante :

- a chaque image on extrait l'imagette qui se trouve juste sous la position du motif sous l'image courante.
- on calcule la différence des intensités entre les deux imagettes (le motif et l'image extraite)
- on remplit le vecteur B

• on multiplie B par $A^{\dagger}$ et on obtient $\begin{bmatrix} u \\ v \end{bmatrix}$.

• on remet a jour la position de la fenêtre définissant le motif et on passe à l'image suivante.

Si on calcule le nombre d'opérations élémentaires que nécessite cette méthode, on s'aperçoit qu'elle est beaucoup plus rapide. Elle est donc avantageuse lorsqu'il faut réduire les temps de calculs. Par contre elle ne peut fonctionner que si l'approximation différentielle est valide, c'est à dire si le mouvement est faible entre les deux images. Inconvénient que ne présentent pas les méthodes de corrélation et de distance.

5• Travail sur machine.

Vous disposez d'un fichier du nom de Poursuite.m qui contient un début de solution utilisant la SAD.
Dans un premier temps vous modifierez ce programme pour n'extraire des imagettes que dans un domaine restreint autour de la position courante.
Dans un deuxième temps, vous ferez évoluer la position courante. Vous devez obtenir un algorithme assez rapide.
Vous testerez ce programme sur les différentes séquences d'images dont vous disposez.
Vous essayerez ensuite d'implanter les autres méthodes pour constater des différences de comportement et de temps de calcul. La plus rapide devrait être la technique du flot optique. C'est aussi celle qui vous demandera la plus de réflexion.
Les séquences nommées Ghost(1 à 7) sont des extraits du film Ghost in the Shell avec des travellings (sauf pour la séquence 7) ... sélectionnez un détail et voyez ce qui se passe.

Remarque générale.

Il vous est conseillé de programmer **proprement** c'est à dire en définissant tous vos paramètres sous forme de variable (avec des noms explicites) et en commentant le plus possible ce que vous écrivez.