

Détection de mouvement

1 Objectifs

Ce projet a 3 objectifs :

- algorithmique : implémenter une chaîne de traitement représentative d'applications en traitement d'images pour les systèmes embarqués.
- optimisation : réaliser un ensemble d'optimisations (algorithmiques, logicielles et architecturales) afin d'accélérer l'exécution de cette chaîne de traitement et comprendre leur apport séparé et combiné.
- méthodologie : développer une méthodologie d'implantation et validation pour améliorer ses capacités de debug.

Ce travail doit être un travail *original* pour chaque binôme. L'utilisation de fragments de code existants est strictement interdite tout comme le partage de code entre binômes.

2 Chaîne de traitements

La chaîne de traitement est composée de deux parties. La première partie est l'algorithme de détection de mouvement qui à partir d'une séquence d'images en niveaux de gris fournit une séquence d'images (de masques) binaires indiquant quels sont les pixels en mouvement (appartenant à une région en mouvement) et quels sont les pixels immobiles (appartenant au fond fixe).

Le problème de la détection du mouvement consiste à séparer dans chaque image les zones en mouvement et les zones statiques. À chaque instant, chaque pixel doit ainsi être étiqueté par un identifiant binaire fixe/mobile. Lorsque la caméra est fixe, on peut effectuer une telle détection à partir des différences temporelles calculées pour chaque pixel.

Soient les notations suivantes :

- t : instant de temps courant, servant à indiquer les images
- I_t : image source en niveau de gris à l'instant t et I_{t-1} : l'image source à l'instant $t - 1$,
- M_t : image de fond (image de moyenne),
- O_t : image de différence, en niveau de gris
- V_t : image de variance (d'écart type en fait) calculée pour tout pixel,
- E_t : image d'étiquettes binaires (mouvement / fond), $E_t(x) = \{0, 1\}$ ou $E_t(x) = \{0, 255\}$ pour coder {fond, mouvement}
- x : le pixel courant de coordonnées (i, j) .

La plupart des techniques de détection du mouvement dans la séquence d'images $I_t(x)$ se fondent sur une estimation du module du gradient temporel $|\frac{\partial I}{\partial t}|$. Si les conditions d'éclairage de la scène varient lentement (sont *constantes* entre deux images consécutives) alors une variation significative du niveau de gris d'un pixel (supérieur à un seuil θ) entre deux images impliquera qu'il y a un mouvement en ce point.

2.1 Algorithme Frame-Difference (FD)

L'algorithme de différence d'images est l'un des premiers algorithmes inventés pour détecter un mouvement. C'est aussi le plus simple : pour un pixel donné, si la différence entre les valeurs de ce pixel pour deux images successives est supérieure à un seuil fixe (et constant dans toute l'image), alors ce pixel est en mouvement. Mais c'est aussi le moins robuste car il fait l'hypothèse que le bruit (la variation du niveau de gris d'un pixel - sans qu'il y ait mouvement) est constant dans toute l'image (c'est pour cela que θ est constant).

2.2 Algorithme Sigma-Delta (SD ou $\Sigma\Delta$)

L'algorithme $\Sigma\Delta$ fait au contraire l'hypothèse que le niveau de bruit (écart type) peut varier en tout point. Pour cela le niveau de gris d'un pixel est modélisé par une moyenne $M_t(x)$ et une variance (un écart type en fait) $V_t(x)$. Si la

Algorithm 1: Frame Difference (FD)

```
1 foreach pixel  $x$  do // step #1 :  $O_t$  computation
2    $O_t(x) = |I_t(x) - I_{t-1}(x)|$ 
3 foreach pixel  $x$  do // step #2 :  $O_t$  thresholding and  $\hat{E}_t$  estimation
4   if  $O_t(x) < \theta$  then  $\hat{E}_t(x) \leftarrow 0$ 
5   else  $\hat{E}_t(x) \leftarrow 1$ 
6
```

différence entre l'image courante et l'image de fond est supérieure à un N fois l'écart type alors il y a mouvement. La valeur de N est un paramètre connu et fixé à l'avance. Typiquement N vaut 2 ou 3, éventuellement 4.

Il s'agit d'une détection de mouvement fondée sur une estimation des statistiques du fond statique basée sur la modulation Σ - Δ : c'est une méthode itérative de conversion analogique/numérique qui incrémente ou décrémente d'une unité la valeur numérisée en fonction du résultat de la comparaison entre la valeur analogique et la valeur numérisée courante.

L'initialisation de l'algorithme pour $t = 0$ est la suivante : $M_0(x) = I_0(x)$ et $V_0(x) = V_{min}$. L'algorithme est appliqué à la séquence d'images à partir de $t = 1$. Les valeurs V_{min} et V_{max} sont des constantes permettant de restreindre les valeurs possibles de V_t . Typiquement, $V_{min} = 1$ et $V_{max} = 254$.

Algorithm 2: $\Sigma\Delta$

```
1 foreach pixel  $x$  do // step #1 :  $M_t$  estimation
2   if  $M_{t-1}(x) < I_t(x)$  then  $M_t(x) \leftarrow M_{t-1}(x) + 1$ 
3   if  $M_{t-1}(x) > I_t(x)$  then  $M_t(x) \leftarrow M_{t-1}(x) - 1$ 
4   otherwise do  $M_t(x) \leftarrow M_{t-1}(x)$ 
5 [step #2 : difference computation]
6 foreach pixel  $x$  do // step #2 :  $O_t$  computation
7    $O_t(x) = |M_t(x) - I_t(x)|$ 
8 foreach pixel  $x$  do // step #3 :  $V_t$  update and clamping
9   if  $V_{t-1}(x) < N \times O_t(x)$  then  $V_t(x) \leftarrow V_{t-1}(x) + 1$ 
10  if  $V_{t-1}(x) > N \times O_t(x)$  then  $V_t(x) \leftarrow V_{t-1}(x) - 1$ 
11  otherwise do  $V_t(x) \leftarrow V_{t-1}(x)$ 
12   $V_t(x) \leftarrow \max(\min(V_t(x), V_{max}), V_{min})$  // clamp to  $[V_{min}, V_{max}]$ 
13 foreach pixel  $x$  do // step #4 :  $\hat{E}_t$  estimation
14   if  $O_t(x) < V_t(x)$  then  $\hat{E}_t(x) \leftarrow 0$ 
15   else  $\hat{E}_t(x) \leftarrow 1$ 
16
```

2.3 Morphologie Mathématique

Dans ce projet nous utiliserons des éléments structurants carrés B de taille 3×3 ou 5×5 . Soit X l'ensemble des pixels de l'image associés à l'élément structurant B . Il existe deux opérations de base : la dilatation de X notée $\delta_B(X)$ et l'érosion de X noté $\epsilon_B(X)$.

L'application des opérateurs de morphologie mathématique est similaires aux opérateurs de filtrage (stencils), mais avec des opérations non linéaires.

Pour des images binaires, la dilatation consiste à calculer un **OR** sur le voisinage B dans l'image source et à l'écrire dans l'image destination. Inversement l'érosion consiste à calculer un **AND** sur le voisinage.

Ainsi, il suffit qu'un point soit à 1 dans le voisinage pour que la dilatation produise un 1 (puisque **x OR 1 = 1**), dilatant ainsi la composante connexe binaire (un ensemble de pixels connexes à 1 dans l'image). Inversement, il suffit qu'un pixel soit à 0 dans le voisinage B pour que l'érosion produise un 0 (puisque **x AND 0 = 0**) érodant ainsi la composante connexe.

L'érosion sert à réduire le bruit dans les images : si l'on considère qu'un petit groupe de pixels est du bruit que l'on cherche à supprimer, alors l'application d'une érosion avec un élément structurant B de taille 3×3 fera disparaître tout groupe de pixels de rayon inférieur à sa taille.

Soit r le rayon et $d = 2r + 1$ le diamètre d'un élément structurant carré B , alors une érosion de rayon r enlève à toute composante connexe une épaisseur de r pixels de contour tandis qu'une dilatation de rayon r ajoute une épaisseur de r pixels au contour (Fig. 2, la logique est inversée : les pixels à 1 sont noirs tandis que les pixel à 0 sont blancs)

Les opérations de morphologie binaire sont étendues aux images en niveaux de gris en remplaçant le **OR** par un *max* et le **AND** par un *min*.

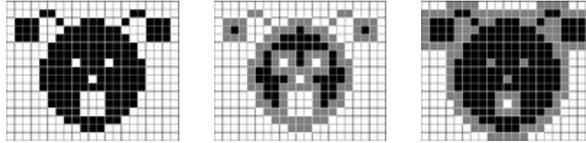


FIGURE 1 – A gauche, image binaire initiale. Au centre, image érodée par un carré 3×3 : les pixels gris sont supprimés de l'ensemble résultant. A droite, image dilatée par un carré 3×3 : les pixels gris sont ajoutés à l'ensemble résultant. Source wikipedia (https://fr.wikipedia.org/wiki/Morphologie_mathmatique)

A partir de ces deux opérateurs, il est possible d'en définir deux autres : la fermeture $\phi_B(X) = \epsilon_B(\delta_B(X))$ et l'ouverture $\gamma_B(X) = \delta_B(\epsilon_B(X))$. La fermeture permet de réduire (voire fermer complètement) les trous à l'intérieur des composantes connexes (si leur rayon est inférieur à celui de l'élément structurant) tandis que l'ouverture fait l'inverse et permet d'agrandir ces mêmes trous.

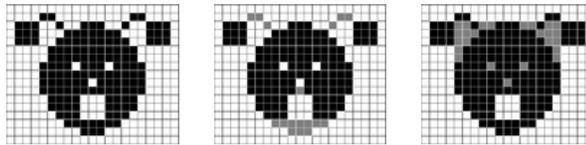


FIGURE 2 – A gauche, image binaire initiale. Au centre, image ouverte par un carré 3×3 : les pixels gris sont supprimés de l'ensemble résultant. A droite, image fermée par un carré 3×3 : les pixels gris sont ajoutés à l'ensemble résultant. Source wikipedia (https://fr.wikipedia.org/wiki/Morphologie_mathmatique)

Un des intérêts de l'ouverture et de la fermeture est qu'elles conservent la taille (discrète) des régions, contrairement à l'érosion qui la diminue ou à la dilatation qui l'augmente. En fonction des besoins, on choisira une fermeture ou une ouverture. Comme ces opérateurs sont idempotents, les appliquer plusieurs fois ne change pas le résultat (qui sera identique à celui obtenu après une seule application). Par contre il est possible de les enchaîner (ouverture puis fermeture ou fermeture puis ouverture) pour améliorer l'image résultat (diminution du bruit, remplissage des trous, ...). En faisant augmenter progressivement leur rayon, on obtient les filtres alternés séquentiel, particulièrement efficaces (et avec une complexité particulièrement élevée) pour supprimer du bruit.

3 Travaux à réaliser

Sont listés l'ensemble des travaux possibles.

3.1 Travail algorithmique

Dans une première étape, il s'agit de réaliser la chaîne de traitement : SD+morpho, en scalaire sans optimisation afin d'obtenir une version de référence.

Dans une seconde étape il s'agit d'optimiser cette chaîne de traitement pour qu'elle soit le plus rapide possible en appliquant des transformations bas niveaux (transformation de boucles) et en appliquant des transformations de haut niveau : pipeline d'opérateurs, fusion d'opérateurs. Le traitement morphologique à implémenter est la séquence érosion $_{3 \times 3}$ - dilatation $_{3 \times 3}$ - dilatation $_{3 \times 3}$ - érosion $_{3 \times 3}$. La séquence d'images à utiliser est `car3.pgm`.

3.2 Travail d'optimisation

Sont listés dans cette section les différentes optimisations possibles (codage SIMD et pipeline ou fusion sont obligatoire). L'évaluation se fera sur la performance du code.

3.2.1 Optimisations *Domain Specific* : fusion / factorisation d'opérateurs

Les opérateurs 2D de morphologie mathématique utilisés sont séparables, factorisables, associatifs et commutatifs :

- Un opérateur avec un élément structurant 2D $d \times d$ est séparable en deux opérateurs 1D d'élément structurant respectif $d \times 1$ et $1 \times d$ et vice-versa. Par exemple (Fig. 3) $(3 \times 3) \leftrightarrow (3 \times 1) \circ (1 \times 3)$ et $(5 \times 5) \leftrightarrow (5 \times 1) \circ (1 \times 5)$.
- Deux opérateurs d'élément structurant carré de rayon r donnent un opérateur d'élément structurant carré de rayon $2r$ et vice-versa. Par exemple (Fig. 4) $(3 \times 3) \circ (3 \times 3) \leftrightarrow (5 \times 5)$.

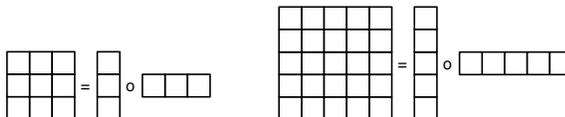


FIGURE 3 – Gauche : décomposition d'un élément structurant (3×3) en deux éléments structurants (3×1) et (1×3) . Droite : décomposition d'un élément structurant (5×5) en deux éléments structurants (5×1) et (1×5) .

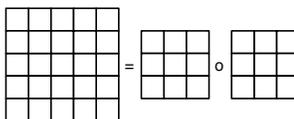


FIGURE 4 – Décomposition d'un élément structurant (5×5) carré de rayon 2 en deux éléments structurants carré (3×3) de rayon 1.

3.2.2 Optimisations logicielles : pipeline d'opérateurs

- Une fois que chaque opérateur de morphologie mathématique a été optimisé avec les transformations précédentes, il est possible d'en optimiser l'enchaînement. Plutôt que d'appliquer un opérateur à une image entière puis de faire de même avec l'opérateur suivant, il est possible de les pipeliner (et de pipeliner la détection avec le post-traitement).
- Les opérateurs sur les images sont décomposés en opérateurs "ligne". Ce sont ces opérateurs "ligne" qui sont pipelinés. Cela permet d'améliorer la persistance des données dans les caches proches du processeur. Cette optimisation est particulièrement efficace dans un contexte multithreadé. On parle alors de *Cache Level Parallelism* qui est une version améliorée du *Thread Level Parallelism* (qui ne s'intéresse qu'aux calculs).
- Cette optimisation est applicable en scalaire, en SWP et en SIMD et peut nécessiter un prologue et un épilogue.

3.2.3 Optimisation des formats de calcul et de stockage en mémoire

- En SIMD et en SWP pour ce qui concerne ce projet, il est primordial de garder le parallélisme maximal pour l'ensemble des calculs.
- Pour la morphologie mathématique binaire, au lieu d'utiliser un octet pour stocker un pixel, il est possible de stocker 8 pixels binaires par octet. On a ainsi 8 pixels dans des matrices `uint8`, 32 pixels dans des matrices `uint32` et 64 pixels dans des matrices `uint64`.
- Enfin, si les données occupent moins de place en mémoire (*8times*1 bit par octet vs *1times*1 bit par octet) elles sont plus rapidement chargées/écrites en mémoire. Cette optimisation permet de réduire le temps de traitement en réduisant les temps d'accès mémoire.

3.2.4 Optimisations matérielles combinées SWP/SIMD \times OpenMP

Le fait de combiner les deux parallélismes matériels que sont le SWP(SIMD) et le multi-coeurs nécessite d'optimiser fortement les accès mémoire car les données vont être consommées à très grande vitesse. Trois possibilités :

- Diminuer le nombre d'accès mémoire en fusionnant des opérateurs. On peut envisager de fusionner des étapes de SD avec les opérateurs "lignes" de morpho-math.
- Maximiser la persistance des données dans les caches pour diminuer la durée des transferts en pipelinant les opérateurs. On peut là aussi envisager de pipeliner des étapes de FD les opérateurs "lignes" de morpho-math.
- Combiner fusion et pipeline. La combinaison "ultime" étant la fusion/pipeline de l'ensemble des opérateurs de morpho 1 bit en SWP/SIMD \times OpenMP.

3.3 Codage

En plus des bibliothèques d'allocation mémoire au format NRC (`nrcutil` contient les routines pour lire et écrire des images au format PGM), un code de base est fourni contenant le chargement et l'écriture des images (`LoadPGM` et `MLoadPGM` pour la lecture de séquence lorsque la taille de chaque image est connue) ainsi que la séquence d'appel aux opérateurs de TI (SD, érosion, dilatation, dilatation, érosion).

3.3.1 Nom et types

Les variables et les tableaux doivent être fortement typés et à la norme NRC. Le nom des variables, des tableaux et des fonctions doit être simple sans être ambigu. Il est recommandé d'utiliser les mêmes noms de tableaux que ceux utilisés dans cet énoncé. Voici quelques exemples possibles :

- `routine_FrameDifference`,
- `routine_FrameDifference.SIMD`,
- `SigmaDelta_step0` pour l'initialisation,
- `SigmaDelta_1step` pour les itérations suivantes.

Il en va de même pour les noms de fichiers (obligatoire) :

- `mouvement.c` contiendra l'ensemble des algorithmes scalaires de détection de mouvement,
- `mouvement.SWPD.c`, pour les implémentations SWP,
- `morpho.c` et `morpho.SWP.c` pour la morphologie mathématique,
- `test_mouvement.c`, `test_mouvement.SWP.c`, `test_morpho.c`, `test_morpho.SWP.c` pour les tests,
- `bench_mouvement.c`, `bench_mouvement.SWP.c`, `bench_morpho.c`, `bench_morpho.SWP.c` pour le benchmarking.

En particulier, pour les tests unitaire SWP, il est obligatoire de valider toutes les versions via des tests unitaires. Les macros permettent de simplifier le passage d'une version à l'autre tout en rendant le debug plus efficace. Leur utilisation est obligatoire.

3.3.2 Codage logique et codage binaire

Il y a deux possibilités pour coder les images de morphologie mathématique. Soit sur $\{0, 1\}$, soit sur $\{0, 255\}$. Le second choix est plus simple car permet une visualisation et une mise-au-point plus rapide. Le premier choix permet de stocker les images booléennes sur 1 bit. Il est conseillé de faire une fonction de conversion d'un format vers l'autre pour avoir les deux à la fois et faire du debug.

3.3.3 Exemple de macro

Afin d'améliorer la sémantique du code, il est conseillé d'utiliser des macros C (ou fonction inline C++ ou wrappers C++) comme

- `load2(T, i, j)`, `store2(T, i, j, x)` pour les accès mémoire
- `or3(x0,x1,x2)`, `or5(x0,x1,x2,x3,x4)` pour la morpho
- `left1(a,b)`, `right1(b,c)` pour les stencils 3×3 et `left2(a,b)`, `right2(b,c)` pour les stencils 5×5

Ces quelques macros permettront de rendre le code portable.

4 Benchmark quantitatif

Il vous revient les choix pour valider et mettre en valeur les optimisations réalisées. Ce peut être des courbes ou des tableaux.

- le temps de traitement en seconde (ou en ms) de chaque opérateur et de la chaîne complète,
- le *cpp* (cycle par point) et le débit (en pixel par seconde)

Il est aussi conseillé de faire varier la taille des images pour analyser les performances des codes optimisés : partir d'images petites tenant dans les caches à des images trop grandes pour tenir dans les caches. Dans ce cas, ce sont le *cpp* et le débit qui sont intéressants à tracer sur une courbe.

La séquence `car3` est disponible à l'adresse suivante :

<http://www-soc.lip6.fr/~lacas/ImageProcessing/MotionDetection/car3.zip>.