

Master Imagerie - UPMC

UE VISION

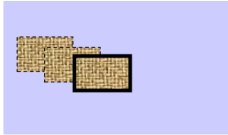
Détection de Mouvement

Antoine Manzanera
ENSTA-ParisTech



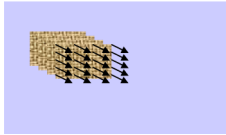
Positionnement dans l'UE VISION

Trois catégories de traitement d'images liés à l'analyse du mouvement :



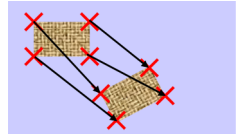
Détection

*Séparer les pixels
des objets mobiles
et du fond statique*



Estimation

*Calculer la vitesse
apparente de
chaque pixel*



Poursuite

*Apparier certaines
structures d'une
image à l'autre*

Contenu et Objectifs du Cours

- Présenter les caractéristiques, enjeux et difficultés de la détection d'objets mobiles dans les séquences d'images.
- Expliquer le principe et développer les différentes techniques de modélisation de fond utilisées en détection de changement temporel.
- Exposer brièvement les méthodes de régularisation spatio-temporelles associées à la détection de changement temporel.

Plan du Cours

- 1 Introduction
 - Contexte et Objectifs
 - Problématique
 - Détection de changement
- 2 Estimation de Fond Statique
 - Moyennes récursives
 - Estimation de densité
 - Estimation Σ - Δ
 - Estimation multi-modale
 - Méthodes par Echantillonnage-Consensus
- 3 Régularisation spatio-temporelle
 - Champs de Markov
 - Morphologie Spatio-temporelle
- 4 Conclusion

Domaines d'application

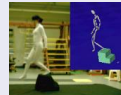
Vidéosurveillance

- Accès
- Comportement
- Reconnaissance



Interaction Homme-Machine

- Commande visuelle
- Contrôle avatar
- Langue des signes



Assistance

- Analyse de la marche
- Aide à la personne
- Analyse sportive



Détection de Mouvement

Contexte

- Caméra fixe
- Environnement non contrôlé

Fond et Objets

Fond (Background) : Scène fixe (mais pas forcément statique !)

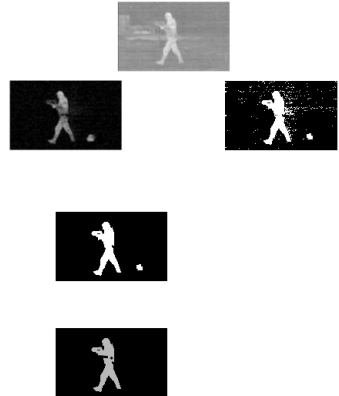
Objets (Foreground) : Objets mobiles (mais pas en permanence !)

- Statistiques temporelles
- Estimation Robuste
- Complexité : Temps et Mémoire



Détection : vue globale

- 1 **Calcul du changement temporel** : Des statistiques temporelles sont calculées pour chaque pixel, d'où l'on peut déduire les valeurs marginales.
- 2 **Régularisation spatio-temporelle** : Les résultats sont agrégés de façon à obtenir des résultats réguliers.
- 3 **Sélection des objets** : Les régions obtenues sont sélectionnées sur des critères morphologiques ou cinématiques.



Quelle observation ?

Quelles variations temporelles faut-il considérer ?

Gradient temporel

$$D_t = |I_t - I_{t-1}|.$$

- ⊕ Très simple
- ⊕ Adaptation simple
- ⊖ Problème de l'ouverture

Valeurs marginales

$$D_t = |I_t - B_t|.$$

- ⊕ Problème de l'ouverture
- ⊕ Gestion des fonds complexes
- ⊖ Adaptation plus difficile

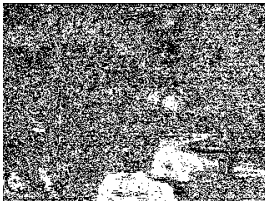
Gradient temporel



I_t (256 niveaux de gris)



$$D_t = |I_t - I_{t-1}|$$



Seuil de D_t à 3



Seuil de D_t à 9

Détermination du seuil

Il est possible d'adapter le niveau *global* du seuil en :

- 1 Supposant que les points isolés sont dûs au bruit.
- 2 Déterminant une proportion cible p_{cible} de points isolés.

Soit p la proportion de points isolés dans l'image binaire.

Si $p < p_{cible}$ alors $\tau_t \leftarrow \tau_{t-1} - 1$, sinon $\tau_t \leftarrow \tau_{t-1} + 1$.



$\tau_t = 2$



$\tau_t = 4$



$\tau_t = 8$



$\tau_t = 15$

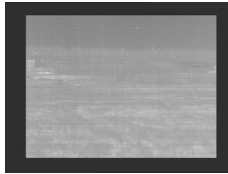


$\tau_t = 25$

Estimation de fond statique



Vidéo



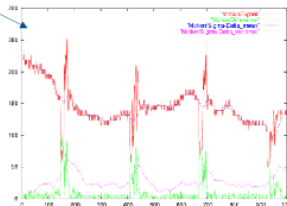
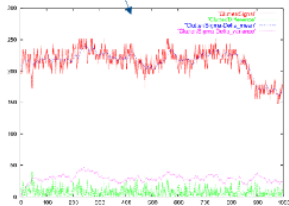
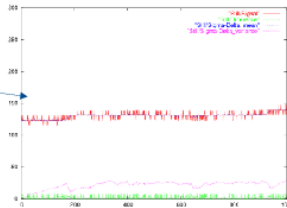
Fond (Background)



Objets (Foreground)

- Traitement de séries temporelles
- Estimation non stationnaire
- Classification Foreground/Background

Un problème d'estimation robuste...



Moyenne temporelle ?

Moyenne récursive naïve

$$B_t = \frac{1}{t} I_t + \frac{t-1}{t} B_{t-1}$$

- Calcul récursif de la moyenne arithmétique
- Pas calculable pour les grandes valeurs de t !

Moyenne temporelle

Filtre exponentiel

$$B_t = \alpha I_t + (1 - \alpha) B_{t-1} ; \alpha \in]0, 1[$$

- α est le paramètre d'oubli (learning rate) ; $\alpha \approx \frac{1}{t}$
- Si $\alpha = 2^{-N}$: calcul très efficace
- Formulation incrémentale $B_t = B_{t-1} + \alpha(I_t - B_{t-1})$

Formulation incrémentale généralisée

Estimation de fond récursive (1er ordre)

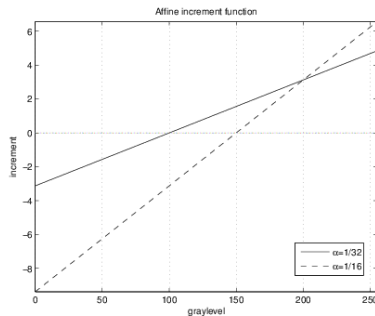
$$B_t = B_{t-1} + \delta_t(I_t, B_{t-1})$$

Pour le filtre exponentiel :

$$\delta_t(I_t, B_{t-1}) = \alpha(I_t - B_{t-1})$$

La fonction d'incrément est linéaire...

Figure : 2 exemples de fonctions d'incrément pour le filtre exponentiel.



Filtre exponentiel bi-niveaux

Moyenne temporelle à deux niveaux

$$B_t = B_{t-1} + \alpha_1(l_t - B_{t-1}) ; \text{ si } l_t \in \text{Fond}$$

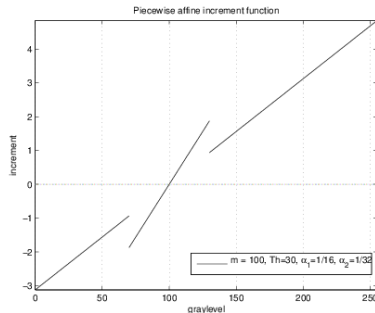
$$B_t = B_{t-1} + \alpha_2(l_t - B_{t-1}) ; \text{ si } l_t \in \text{Objets } (\alpha_2 \ll \alpha_1)$$

Un critère de classification est nécessaire.

P. ex, seuillage :

$$|l_t - B_{t-1}| > \tau_t$$

Figure : 1 exemple de fonctions d'incrément pour le filtre exponentiel bi-niveaux.



Estimation récursive des moyenne et variance

Le même schéma récursif peut être utilisé pour estimer la variance temporelle, permettant d'ajuster localement le seuil de classification Foreground/Background :

Moyenne et Variance Récursives

$$D_t = I_t - B_{t-1}$$

Si $|D_t| > n\sqrt{V_{t-1}}$, $E_t = 1$ (Foreground), sinon $E_t = 0$ (Background).

$$B_t = B_{t-1} + \alpha_t D_t$$

$$V_t = V_{t-1} + \alpha_t D_t^2$$

- B_t est la moyenne, V_t la variance.
- n est un entier, typiquement 2 ou 3.
- $\alpha_t = \alpha_1$ si $E_t = 0$, et $\alpha_t = \alpha_2$ sinon ($\alpha_2 \ll \alpha_1$).

Estimation récursive des moyenne et variance

Moyenne et Variance Récursives

$$D_t = I_t - B_{t-1}$$

Si $|D_t| > n\sqrt{V_{t-1}}$, $E_t = 1$ (Foreground), sinon $E_t = 0$ (Background).

$$B_t = B_{t-1} + \alpha_t D_t$$

$$V_t = V_{t-1} + \alpha_t D_t^2$$

L'estimation de la variance permet d'adapter le seuil localement, mais la fonction d'incrément reste peu satisfaisante car linéaire (α) et/ou discontinue ($\alpha_2 < \alpha_1$).

Estimation pondérée par la densité

En fait, si l'on reprend l'expression incrémentale $B_t = B_{t-1} + \delta_t(l_t, B_{t-1})$, la fonction d'incrément δ_t devrait aussi dépendre de la probabilité d'observer la valeur l_t :

Estimation pondérée (cas général)

$$\delta_t(l_t, B_{t-1}) = \frac{\alpha_{\max} f_t(l_t)}{f_t(B_{t-1})} \times (l_t - B_{t-1})$$

avec :

- $f_t(x) = P(B_t = x)$ densité de probabilité du fond.
- α_{\max} paramètre d'oubli maximal.
- B_{t-1} correspond au mode courant de la distribution.

Estimation de densité temporelle

La densité temporelle peut être estimée en utilisant la mise à jour récursive de l'histogramme :

Algorithme d'estimation de densité

- Soit $\{1, \dots, N\}$ le domaine de l'histogramme (niveaux de gris).
- Initialisation : $f_0(i) = 1/N$ pour tout $i \in \{1, \dots, N\}$
- Pour $t > 0$:
 - $f_t(I_t) = f_{t-1}(I_t) + \varepsilon$
 - Renormaliser f_t

La valeur de référence du fond B_t (si nécessaire) peut être définie comme le *mode* de l'histogramme $\arg \max_{i \in \{1, \dots, N\}} f_t(i)$, ou comme la *médiane* $F_t^{-1}(1/2)$, avec $F_t(i) = \sum_{j < i} f_t(j)$.

Estimation de densité temporelle

Algorithme d'estimation de densité

- Soit $\{1, \dots, N\}$ le domaine de l'histogramme (niveaux de gris).
- Initialisation : $f_0(i) = 1/N$ pour tout $i \in \{1, \dots, N\}$
- Pour $t > 0$:
 - $f_t(I_t) = f_{t-1}(I_t) + \varepsilon$
 - Renormaliser f_t

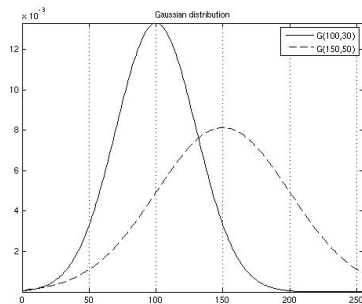
La classification peut aussi être faite directement (i.e. sans estimer de fond B_t) à partir de la densité : par exemple $E_t = 1$ si $f_t(I_t) < \tau$.

Estimation de densité gaussienne

Si la densité correspond à un modèle connu, l'estimation peut être simplifiée, par exemple dans le cas de la gaussienne simple (1 mode/moyenne, 1 variance) :

Distribution gaussienne

$$f_t(x) = \frac{1}{\sigma_t \sqrt{2\pi}} \exp \left(-\frac{(x - \mu_t)^2}{2\sigma_t^2} \right)$$



Estimation de densité gaussienne

Fonction d'incrément gaussienne

$$\delta_t(I_t, B_{t-1}) = \alpha_{max} \times \exp\left(\frac{-(I_t - B_{t-1})^2}{2V_{t-1}}\right) \times (I_t - B_{t-1})$$

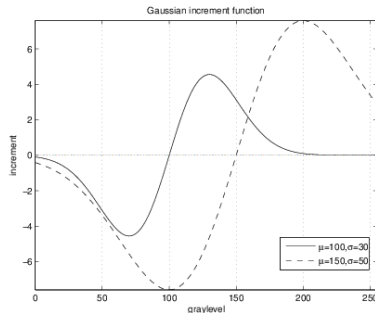
Estimation de la variance :

$$V_t = V_{t-1} + \alpha_V((I_t - B_t)^2 - V_{t-1})$$

Classification :

$$E_t = 1 \Leftrightarrow |I_t - B_t| > k \times \sqrt{V_t}$$

Figure : 2 exemples de fonctions d'incrément pour l'estimation de densité gaussienne.

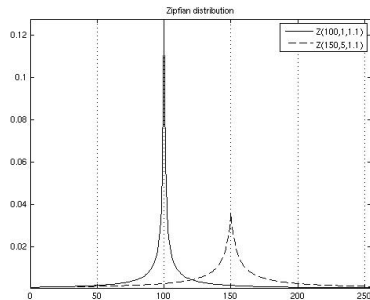


La distribution de Zipf-Mandelbrot

Distribution Zipfienne centrée

$$Z_{(\mu,k,s)}(x) = \frac{(s-1)k^{s-1}}{2(|x-\mu|+k)^s}$$

- μ est la moyenne (mode) de la distribution
- k détermine la dispersion (\simeq variance)
- $s \simeq 1; s > 1$



La distribution de Zipf-Mandelbrot

Distribution Zipfienne centrée

$$Z_{(\mu,k,s)}(x) = \frac{(s-1)k^{s-1}}{2(|x-\mu|+k)^s}$$

- Origine : linguistique (fréquence des mots dans une langue naturelle).
- Déjà utilisé en traitement d'images spatial (codage, segmentation).
- Utilisé ici comme modèle de distribution temporelle.

Estimation de fond zipfienne

Fonction d'incrément zipfienne

La fonction d'incrément zipfienne peut être approximée par une fonction de Heaviside :

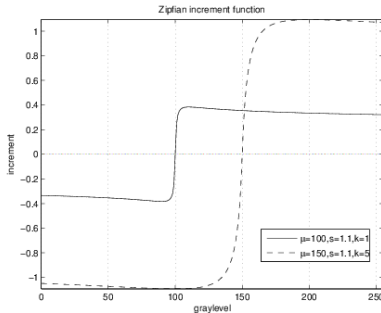
$$\delta_t \simeq H_{(\mu, \kappa)}(x) = -\kappa \text{ si } x < \mu, +\kappa \text{ si } x > \mu \text{ (avec } \kappa = \alpha_{\max} k^s \text{)}$$

Ainsi l'estimation zipfienne peut être approximée par la modulation Σ - Δ :

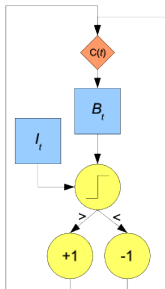
$$B_t = B_{t-1} + \varepsilon \text{ si } I_t > B_{t-1}$$

$$B_t = B_{t-1} - \varepsilon \text{ si } I_t < B_{t-1}$$

L'incrément élémentaire ε devrait dépendre de la variance du fond.
 Figure : 2 exemples de fonctions d'incrément pour l'estimation de densité zipfienne.



L'algorithme d'estimation Σ - Δ (1)

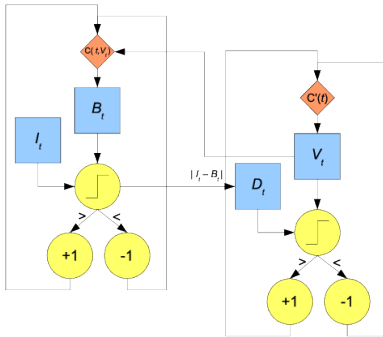


L'incrément élémentaire correspond au bit de poids faible (LSB), i.e. ± 1 .

L'incrément moyen est ajusté temporellement en changeant la fréquence de mise à jour :

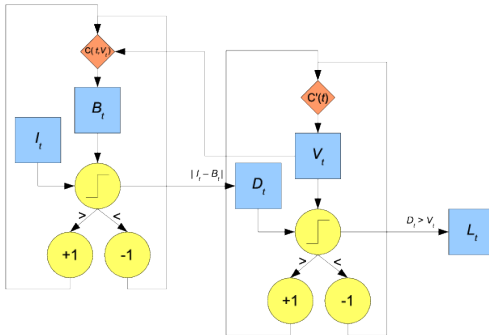
Cela correspond à la condition $C(t)$ (typiquement $C(t) \equiv (t \% n) == 0$)

L'algorithme d'estimation Σ - Δ (2)



Comme l'incrément moyen doit dépendre de la variance du fond, la condition de mise à jour doit aussi dépendre de l'estimateur de dispersion V_t . (Plus V_t est grand, plus la mise à jour doit être fréquente). L'estimateur de dispersion V_t est calculé par estimation Σ - Δ de la séquences des différences absolues $|I_t - B_t|$.

L'algorithme d'estimation Σ - Δ (3)



Finalement, la classe Foreground/Background est obtenue en comparant la différence absolue à l'estimateur de dispersion.

Exemple : Vidéo avec mouvement radial



Original

Background



Variance



Foreground

Évaluation quantitative

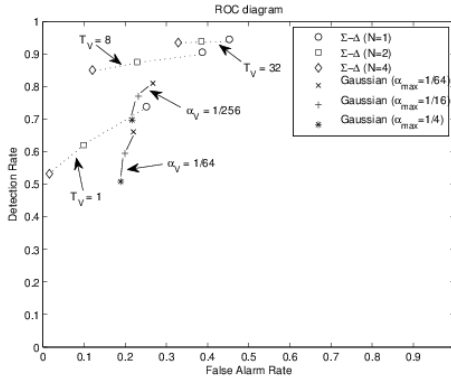


Figure: Comparaison de plusieurs algorithmes de soustraction de fond Σ - Δ et Estimation Gaussienne, avec différents paramètres temporels.

Avantages calculatoires

- Le coût de calcul de l'algorithme Σ - Δ est extrêmement bas :
 - *Mémoire* : 2 descripteurs entiers.
 - *Jeu d'Instruction* : réduit à différence, comparaison, et incrément/décrément.
 - *Taille de données* : Aucune approximation, adapté à l'arithmétique fixe de n'importe quelle taille.
- Il a été implanté sur diverses plate-formes temps réel, dont :
 - *Parallélisme cellulaire* : Rétine programmable PVLSAR 34.
 - *Parallélisme vectoriel*: Extensions multimedia SSE2, Altiivec.
 - *Composants programmables* : FPGA Xilinx XSA3S1000.

Estimation multi-modale de fond

L'utilisation de distributions unimodales comme modèle probabiliste peut être insuffisante dans le cas de fonds complexes (ex : houle, drapeau,...). Mais les méthodes précédentes peuvent être étendues à des modèles multi-modaux (mélanges), comme suit :

Algorithme d'estimation multi-modale

Soit $\{B^i, V^i, W^i\}_{i=1..N}$ les N modes

Pour tout pixel :

pour chaque mode i :

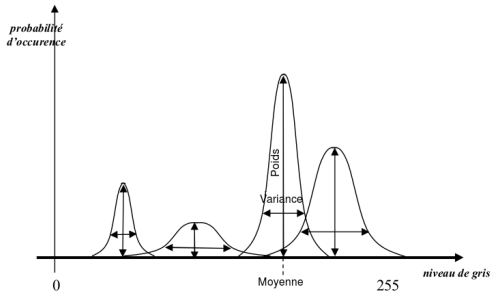
Si $|I_t - B_t^i| < n\sqrt{V_t^i}$:

Mettre à jour le mode correspondant $\{B_t^i, V_t^i, W_t^i\}$ (B^i, V^i est mis à jour comme dans le cas unimodal, W^i est incrémenté puis normalisé)

Trier les différents modes par "importance" $W^i/\sqrt{V^i}$, et choisir le (les) premiers pour représenter le fond.

Estimation multi-modale de fond

- La distribution multimodale est représentée par $3N$ nombres $\{B^i, V^i, W^i\}_{i=1..N}$.
- N est le nombre de modes, typiquement entre 3 et 7.
- B^i et V^i représentent les moyenne (mode) et variance de chaque sous-distribution.
- W^i représentent les poids relatifs des différents modes.

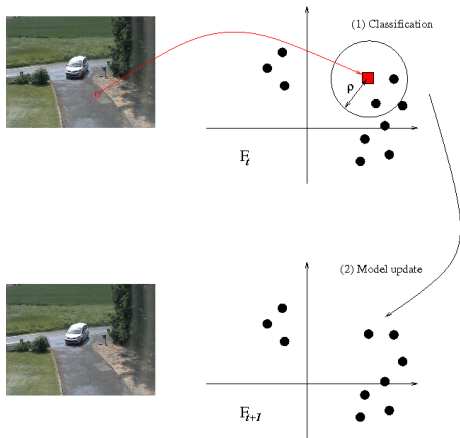


Méthodes par Echantillonnage-Consensus

- Certaines méthodes représentent le fond sans calculer spécifiquement de statistiques, mais simplement en conservant en mémoire quelques valeurs $\{I_{t_1}, \dots, I_{t_K}\}$ (échantillonnage).
- La classification Foreground/Background est ensuite réalisée en jugeant si la valeur courante est proche de l'échantillon (consensus). Exemple ViBe :
$$E_t = 1 \Leftrightarrow |\{i \in \{1, \dots, K\}; d(I_t, I_{t_i}) > \tau\}| > T.$$
- L'échantillon est ensuite mis à jour, en prenant éventuellement en compte la valeur de E_t .

Méthodes par Echantillonnage-Consensus

Les méthodes d'échantillonnage/consensus peuvent s'appliquer sur le niveau de gris, ou sur des espaces multidimensionnels de couleurs ou même de caractéristiques locales.



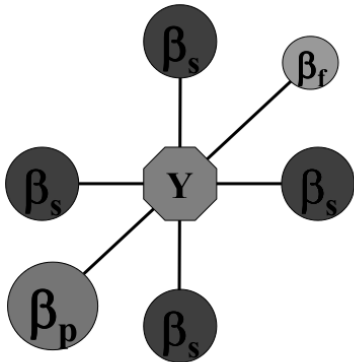
Régularisation par Champs de Markov

La détection de changement temporelle n'est pas suffisante en général. La régularisation spatio-temporelle par champ de Markov a été souvent employée pour la détection de mouvement :

- **Modélisation** : le champ d'étiquette Fixe/Mobile est un champ de Markov dans l'espace-temps discret.
- **Théorème de Hammersley et Clifford** : la densité de probabilité se calcule à partir d'une fonction (énergie) calculée sur les cliques de la maille discrète.
- **Simulation** : on peut obtenir des tirages de ce champ aléatoire (ex: échantillonneur de Gibbs).
- **Optimisation** : on cherche à obtenir la valeur la plus probable du champ (ex: ICM, Recuit simulé).

Voir http://www.ensta.fr/~manzaner/Download/Tutorials/ETASM06_chap1.pdf.

Régularisation par Champs de Markov



$U_m(x)$: énergie du modèle

$U_a(y,x)$: énergie d'attache
aux données

$$U(x) = U_m(x) + U_a(x, y)$$

$$U_m(x) = \sum_{s \in S} \sum_{r \in \mathcal{V}(s)} V_x(s, r)$$

$$\text{avec } V_x(s, r) = -\beta_{sr} \text{ si } x(s) = x(r) \\ = +\beta_{sr} \text{ si } x(s) \neq x(r)$$

(Modèle de Potts)

$$U_a(x, y) = \frac{1}{2\sigma^2} \sum_{s \in S} (y(s) - \Psi(s))^2$$

$$\text{avec } \Psi(s) = 0 \text{ si } x(s) = 0 \\ = \alpha \text{ si } x(s) = 1$$

Régularisation par Champs de Markov

$$U(x) = U_m(x) + U_a(x, y)$$

$$U_m(x) = \sum_{s \in S} \sum_{r \in \mathcal{V}(s)} \pm \beta_{sr}$$

$$U_a(x, y) = \frac{1}{2\sigma^2} \sum_{s \in S} (y(s) - \Psi(s))^2$$

$$\text{avec } \Psi(s) = 0 \text{ si } x(s) = 0 \\ = \alpha \text{ si } x(s) = 1$$

Modélisation :

On suppose :

$$P(X = x) = \frac{e^{-U_m(x)}}{Z_1}$$

(caractère markovien de X)

L'énergie du modèle exprime une hypothèse de régularité.

$$\text{et : } P(Y = y / X = x) = \frac{e^{-U_a(x, y)}}{Z_2}$$

Avec: α = moyenne de Y ; σ^2 = variance de Y

(modèle de bruit liant X et Y)

L'énergie d'adéquation assure un lien significatif entre le résultat de l'étiquetage et les données d'entrée.

$$\text{Alors : } \arg \min_x U(x) = \arg \max_x P(X = x) \cdot P(Y = y / X = x)$$

P. Bouthémy
P. Lalande
1993

$$= \arg \max_x P(X = x / Y = y)$$



Critère bayésien du Maximum A Posteriori

Régularisation par Morphologie Spatio-temporelle

La régularisation spatio-temporelle est souvent réalisée sur les images binaires de Foreground en utilisant les opérateurs morphologiques :

- Filtres alternés séquentiels :

$$F_n(E_t) = \delta_{B_n}(\varepsilon_{B_n}(\delta_{B_{n-1}}(\varepsilon_{B_{n-1}}(\dots \delta_{B_1}(\varepsilon_{B_1}(E_t)) \dots))).$$



E_t



$F_2(E_t)$

Régularisation par Morphologie Spatio-temporelle

Opérateurs morphologiques connexes :

- F.A.S par reconstruction : $E'_t = R_{E_t}(F_n(E_t))$.



E_t



$F_2(E_t)$



E'_t

Régularisation par Morphologie Spatio-temporelle

Opérateurs spatio-temporels connexes :

- Filtre spatio-temporel connexe :

$$E_t'' = R_{E_t} (F_n(E_t) \cap \delta_{B_m}(E_{t-1}'))$$



E'_{t-1}



$\delta_{B_m}(E'_{t-1})$



E'_t



E''_t

Notions clefs à retenir

- Détection de changement \Leftrightarrow Détection des singularités dans les séries temporelles.
- Représentations du fond :
 - Paramètres d'une distribution uni- ou multimodale.
 - Histogramme d'une distribution quelconque.
 - Echantillonnage d'une distribution quelconque.
- Compromis coût de calcul (temps, mémoire) / complexité de la représentation (longueurs et nombres de statistiques / classes de valeurs / modes / échantillons / ...)
- Régularisation spatio-temporelle : Champs de Markov, Morphologie mathématique,...

Références



[Elga99] A. ELGAMMAL, D. HARDWOOD and L.S. DAVIS
Non-parametric Model for Background Subtraction
Proc. of ICCV '99 FRAME-RATE Workshop(1999)



[Stauf00] C. STAUFFER and C. GRIMSON
Learning patterns of activity using real-time tracking.
IEEE Trans. on PAMI 22(8), 747-757. (2000)



[Mittal04] A. MITTAL and N. PARAGIOS
Motion-based background subtraction using adaptive kernel density estimation.
IEEE CVPR'04

Références



[Power02] P. POWER and J. SCHONEES

Understanding background mixture models for foreground segmentation.

In: Imaging and Vision Computing New Zealand, Auckland, NZ (2002)



[Manza07a] A. MANZANERA and J. RICHEFEU

A new motion detection algorithm based on Sigma-Delta background estimation.

Pattern Recognition Letters 28(3), 320-328. (2007)



[Manza07b] A. MANZANERA

Sigma-Delta Background Subtraction and the Zipf Law.

Progress in Pattern Recognition, Image Analysis and Applications (CIARP'07) pp. 42-51.

Références

 **[Wang07]** H. WANG and D. SUTER

A consensus-based method for tracking: Modelling background scenario and foreground appearance
Pattern Recognition 40(3), 1091-1105. (2007)

 **[Barnich09]** O. BARNICH and M. VAN DROGENBROECK

ViBe: a powerful random technique to estimate the background in video sequences
International Conference on Acoustics, Speech, and Signal Processing 945-948. (2009)

 **[Bouthémy93]** P. BOUTHÉMY and P. LALANDE

Recovery of moving object masks in an image sequence using local spatiotemporal contextual information.
Optical Engineering, 32(6):1205-1212, June 1993.