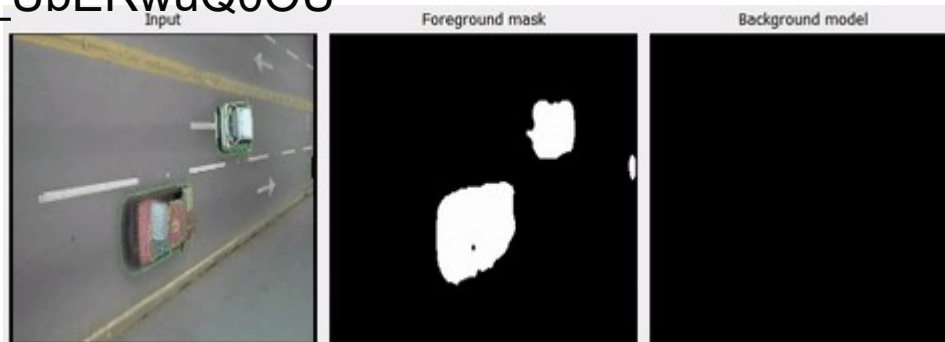


Détection d'Objets en Mouvement par Suppression du Fond

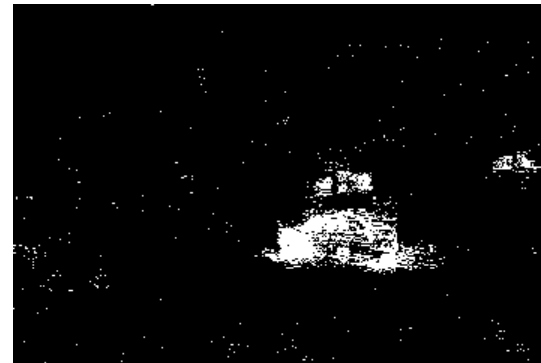
**Inspiré des supports de cours de
Thierry Bouwmans, entre autres ;)**

https://www.youtube.com/watch?v=_UbERwuQ0OU



Objectif

Détection d'objets en mouvement dans une vidéo.



Aucunes connaissances a priori sur le fond et la forme des objets
(voitures, personnes, animaux etc.)

Hypothèse à discuter !!

Applications - Domaine Visible

Video Surveillance - PETS 2006



La vidéosurveillance des lieux publics extérieurs dans l'objectif de détecter des activités anormales

Applications - Domaine Visible

Video Surveillance



La vidéosurveillance de scènes intérieures pour détecter les comportements de vols ou pour optimiser un service en comptant le nombre de personnes par zone ou par rayon dans les magasins pour assurer une bonne répartition des vendeurs selon le besoin

Applications - Domaine Visible

- Video Surveillance



- La vidéosurveillance d'activité portuaire pour compter le nombre de navires

Applications - Domaine Visible

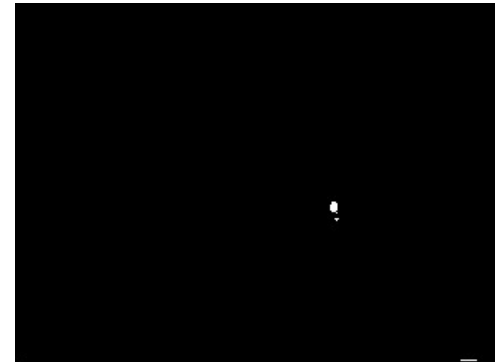
Video Surveillance



La surveillance du trafic routier pour évaluer le nombre de véhicules ou pour détecter les ralentissements ou les accidents (Projet ATON).

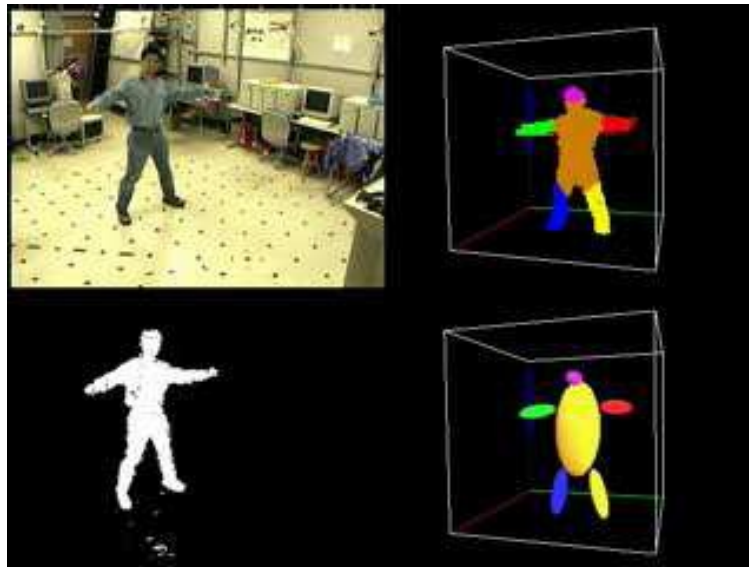
Applications - Domaine Infra Rouge

Video Surveillance - Terravic



Applications - Capture du Mouvement

Capture de Mouvement



Une bibliothèque avec de nombreux algorithmes :

<https://github.com/andrewssobral/bgslibrary>

Les Hypothèses (Fortes)

Hypothèse 1 : La caméra est fixe et ses caractéristiques le sont également.

Hypothèse 2 : La scène ne comporte pas de variations lumineuses.

Hypothèse 3 : Le fond est visible et statique, c'est-à-dire que les pixels ont une distribution unimodale et qu'il n'y a pas d'objet qui bouge ou qui soit inséré dans la scène au cours du temps.

Hypothèse 4 : Le fond ne contient pas d'objets d'intérêt statiques.

Hypothèse 5 : les objets d'intérêt peuvent être facilement détectés en seuillant la différence entre l'image courante et l'image du fond.

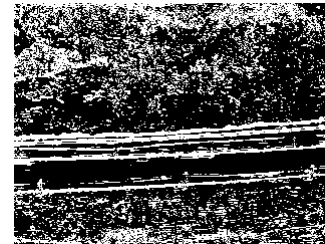
Les Situations Critiques

CS 1 - Image bruitée due à une mauvaise qualité de la caméra.

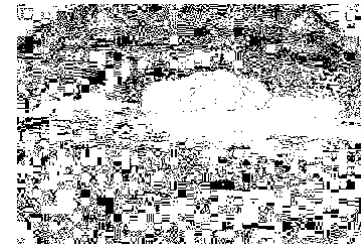


Les Situations Critiques

CS 2-1 - Mouvement de la caméra (Vent).

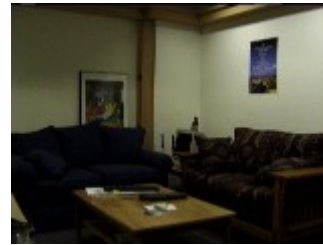
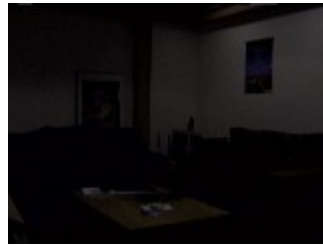


CS 2-2 - Ajustement automatique de la caméra (auto focus, auto white balance, auto exposure correction)

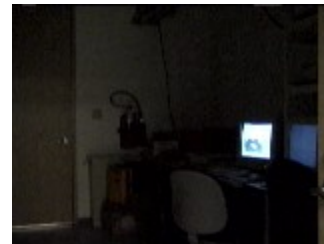


Les Situations Critiques

CS 3 – Changement lumineux lent (Lumière du jour).

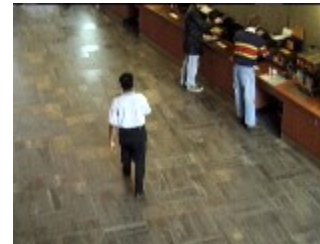


CS 4 - Changement lumineux rapide (Lumière artificielle on/off)



Les Situations Critiques

CS 5 – Bootstrapping: Durant la période d'initialisation, le fond n'est pas visible ou très peu.

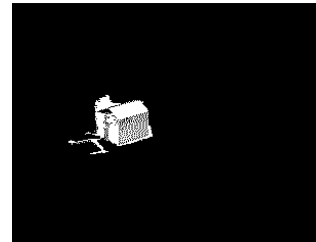


CS 6 – Camouflage: Un objet en mouvement a la même apparence que le fond.



Les Situations Critiques

CS 9 – Insertion d'un objet du fond – Un objet peut être inséré dans la scène. Cet objet ne doit pas être considéré comme un objet en mouvement.



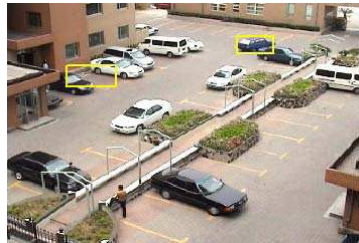
Les Situations Critiques

CS 10 – Mouvement dans le fond: Le fond peut présenter des objets qui ont un mouvement et donc demande un modèle qui peut représenter un ensemble disjoint de pixels.



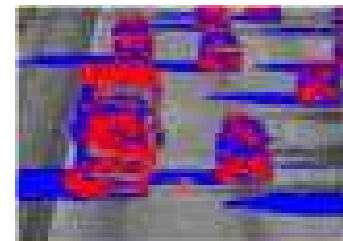
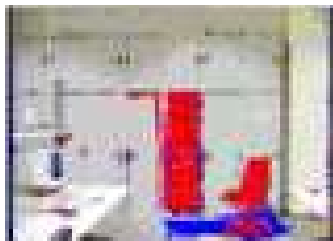
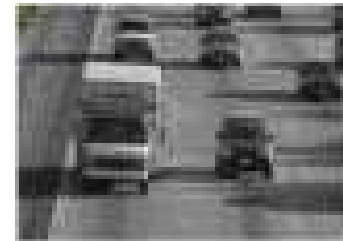
Les Situations Critiques

CS 11 – Objet qui commence à se déplacer: Quand un objet commence à se déplacer, à la fois l'objet et la partie du fond révélée sont détectés.



Les Situations Critiques

CS 13 – Ombres: Les ombres ne doivent pas être détectées comme objet en mouvement.



Les Situations Critiques

Différentes caractéristiques spatiales et temporelles.

Exemple des variations lumineuses

Caractéristiques Spatiales: Variations lumineuses peuvent être locales ou globales.

Caractéristiques Temporelles: Variations lumineuses peuvent être lentes ou rapides.

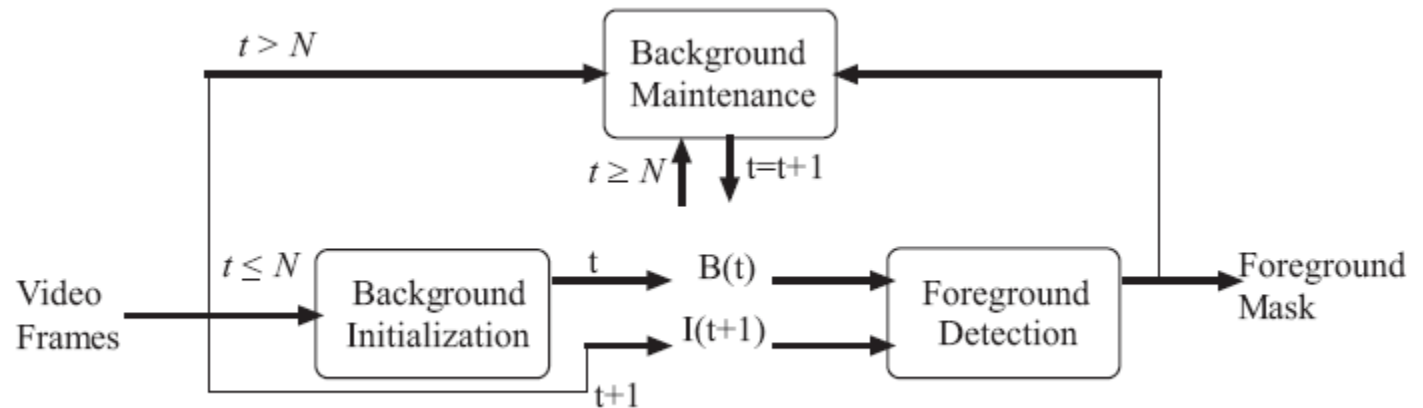
La méthode idéale doit prendre en compte toutes les situations critiques et chacune de leurs caractéristiques spatiales et temporelles.

Les Différentes Etapes et Choix Clés

Représentation du fond

- Initialisation du fond
- Mise à jour du fond
 - Classification objet en mouvement et fond
 - Choix de la taille de l'élément image
 - Choix de la caractéristique

Les Différentes Etapes



Modèle de l'Initialisation du Fond

L'initialisation du fond détermine la première image avec laquelle on commence le processus

Soit la première image courante

Soit une déterminée par un modèle de représentation

Modèle de Mise à Jour

- La mise à jour du modèle permet de prendre en compte les changements apparaissant dans la scène.
- Deux mécanismes sont possibles:
 - Mise à jour non selective: Tous les pixels sont mis à jour avec la même règle.
 - Mise à jour selective: Les pixels classés comme actif ou inactif sont mis à jour avec une règle différente.

Modèle de Mise à Jour – Non Selectif

Tous les pixels sont mis à jour avec la même règle.
Couramment, une moyenne est utilisée:

$$B_{n+1}(x, y) = (1 - \alpha) B_n(x, y) + \alpha I_n(x, y)$$

Où α est le coefficient d'apprentissage

Le choix du coefficient d'apprentissage détermine la vitesse de mise à jour du fond.

Modèle de Mise à Jour – Selectif

- La mise à jour est effectuée en utilisant deux règles différentes:

Si le pixel a été classé comme actif

$$B_{n+1}(x, y) = (1 - \alpha) B_n(x, y) + \alpha I_n(x, y)$$

Si le pixel a été classé comme passif

$$B_{n+1}(x, y) = (1 - \beta) B_n(x, y) + \beta I_n(x, y)$$

- L'idée est de mettre à jour rapidement pour les pixels actifs et lentement les pixels inactifs. Pour cela, $\beta \ll \alpha$

Classification

La classification consiste à décider si le pixel est classé comme Fond ou Objet en mouvement.

Classification - Méthodes de Bases

Les pixels sont classés comme objet en mouvement:

$$|B(x) - I(x)| > T$$

sinon les pixels sont classés comme fond.

Classification

■ Classification - Méthodes Statistiques

Les pixels sont classés comme objet en mouvement:

$$|\mu(x) - I(x)| > k\sigma(x)$$

sinon les pixels sont classés comme fond.

Les pixels sont classés comme objet en mouvement:

sinon les pixels sont classés comme fond.

$$P(x) > T$$

Détection d'Objets en Mouvement par Suppression du Fond

Partie 2: Méthodes de Bases

Thierry Bouwmans

Méthodes de Bases

Dans la littérature, il existe trois grands types de méthodes de bases:

- **Moyenne, Moyenne mobile [Lee 2002]**
- Médiane, Médiane approximée [Mc Farlane 1995]
- Histogramme, Fréquence d'apparition, Intervalle stable [Zheng 2006]

Moyenne – Moyenne mobile

Une façon simple pour obtenir le fond est de calculer la moyenne temporelle:

$$B(x, y, t) = \frac{1}{t} \sum_{t'=0}^{t'} I(x, y, t')$$

ou de façon récursive,

$$B(x, y, t) = \frac{1}{t} B(x, y, t-1) + \left(1 - \frac{1}{t}\right) I(x, y, t)$$

Inconvénients: La moyenne à long-terme peut ne pas capturer correctement les variations lumineuses.

Moyenne – Moyenne mobile

Une meilleure façon de capturer un fond sensible aux variations lumineuses lentes est d'utiliser un facteur d'oubli exponentiel:

$$B(x, y, t) = e^{-1/T} B(x, y, t-1) + (1 - e^{-1/T}) I(x, y, t)$$

où T est la période d'observation.

Cela revient à calculer une moyenne courante pondérée:

$$B(x, y, t) = \alpha B(x, y, t-1) + (1 - \alpha) I(x, y, t)$$

où α est appelé facteur d'oubli exponentiel ou le facteur d'apprentissage. Sa valeur est comprise entre 0 et 1.

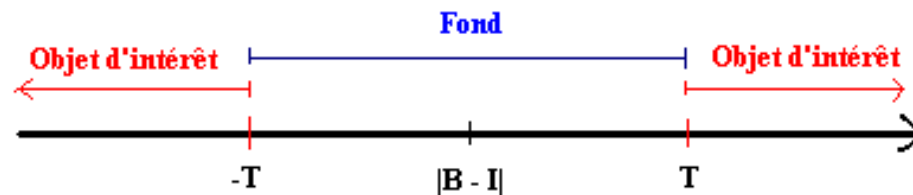
Moyenne – Moyenne mobile

Classification Fond/Objet en mouvement

Un pixel est considéré comme fond si

$$|I_{t+1}(x, y) - B_t(x, y)| < T$$

Sinon le pixel est classé en objet en mouvement



Moyenne – Moyenne mobile

Avantages: Facile à calculer

Inconvénients: Bien que cette méthode soit efficace quand les objets bougent de façon continue, elle n'est pas robuste quand les objets bougent lentement car dans ce cas, ils sont intégrés dans le fond.

Médiane – Médiane approximée

Une façon plus robuste de modéliser le fond est d'utiliser la médiane:

$$B(x, y, t) = \underset{t' \in \{0, t\}}{\text{median}}(I(x, y, t'))$$

Utiliser la médiane présente deux inconvénients majeurs:

- La médiane doit être calculé à chaque instant (Temps de calcul)
- Toutes les frames doivent être stockées (Espace mémoire)

Médiane – Médiane approximée

- Médiane approximée [Schofield 1995] :

$$B(x, y, t+1) = B(x, y, t) + 1 \quad \text{if} \quad B(x, y, t) < I(x, y, t)$$

$$B(x, y, t+1) = B(x, y, t) - 1 \quad \text{if} \quad B(x, y, t) > I(x, y, t)$$

- Autres médianes approximées plus efficaces [Joubert 2004]:
 - Exponentially Weighted Moving Average (EWMA)
 - Stochastic Approximation (SA),
 - Exponentially Weighted Stochastic Approximation (EWSA)
 - New – Collapse – Output (NCO)

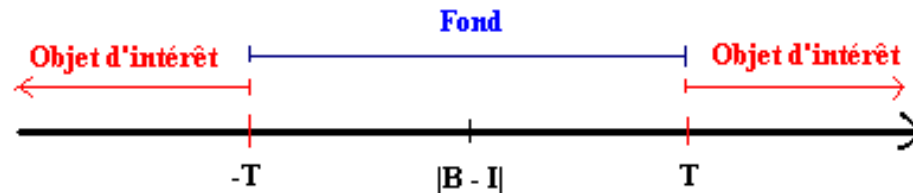
Médiane – Médiane approximée

Classification Fond/Objet en mouvement

Un pixel est considéré comme fond si

$$|I_{t+1}(x, y) - B_t(x, y)| < T$$

Sinon le pixel est classé en objet en mouvement



Méthodes de Bases

Dans la littérature, il existe trois grands types de méthodes de bases:

- Moyenne, Moyenne courante [Lee 2002]
- Médiane, Médiane approximée [Mc Farlane 1995]
- **Histogramme, Fréquence d'apparition, Intervalle stable [Zheng 2006]**

Détection d'Objets en Mouvement par Suppression du Fond

Partie 3: Modèles Statistiques – Première Génération

Thierry Bouwmans

Modèles Statistiques: Une Gaussienne

Une Seule Gaussienne [Wren 1997]

Initialisation du fond: Oui

Représentation du Fond: Une seule Gaussienne

Mise à jour du fond: Oui

Espace Couleur : YUV

Modélisation du Fond

[Wren 1997] a développé un algorithme qui modélise un pixel à partir de loi de distribution supposée normale et caractérisée par sa moyenne et sa variance dans l'espace couleur YUV.

Modélisation par une Gaussienne

Initialisation du Fond

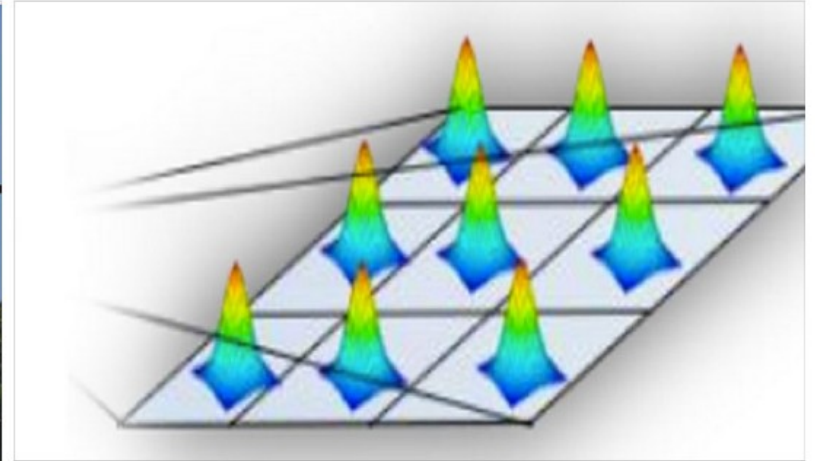
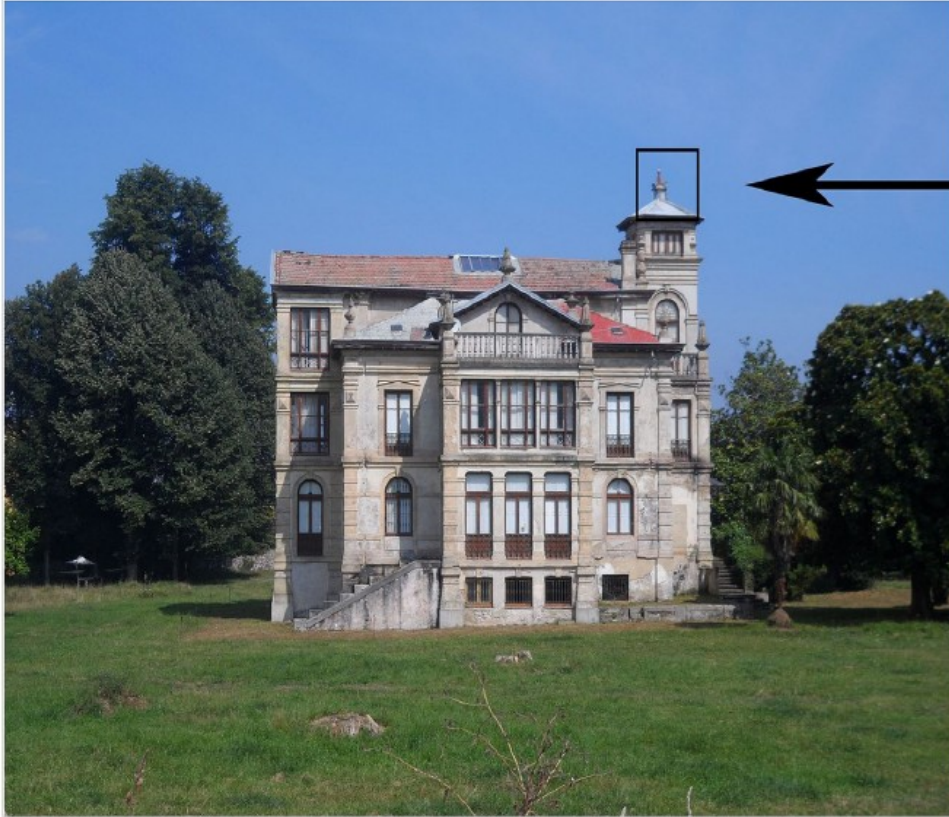
Ce modèle nécessite un nombre t de frames pour calculer la moyenne μ et la variance dans chaque plan couleur:

$$\mu(x, y, t) = \sum_{i=1}^t \frac{I(x, y, i)}{t}$$

$$\sigma(x, y, t) = \sqrt{\left(\sum_{i=1}^t \frac{I^2(x, y, i)}{t} - \mu^2(x, y, t) \right)}$$

Où $I(x, y, i)$ est l'intensité courante du pixel à la position (x, y) au temps i .





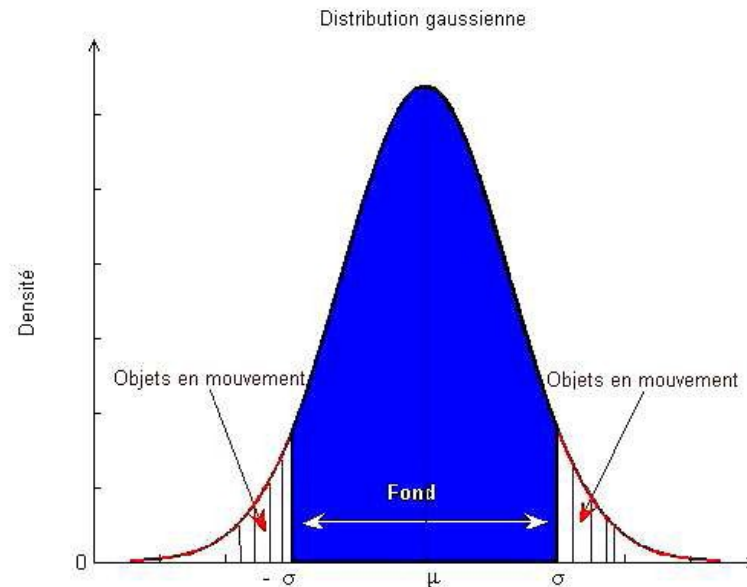
Modélisation par une Gaussienne

Classification

Un pixel est classé comme objet en mouvement si :

$$|\mu(x, y, t) - I(x, y, t)| > c \cdot \sigma(x, y, t)$$

où c est une constante fixe.



Modélisation par une Gaussienne

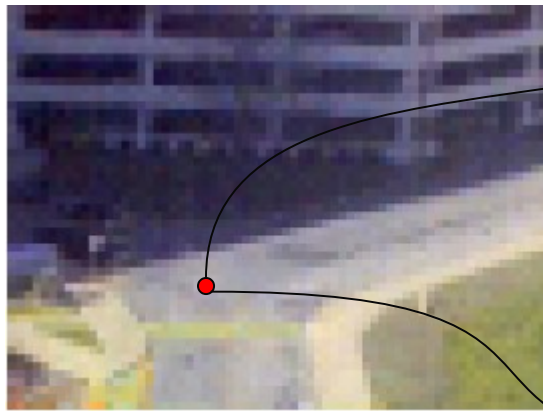
Mise à Jour du Fond

$$\mu(x, y, t+1) = (1 - \alpha)\mu(x, y, t) + \alpha I(x, y, t+1)$$

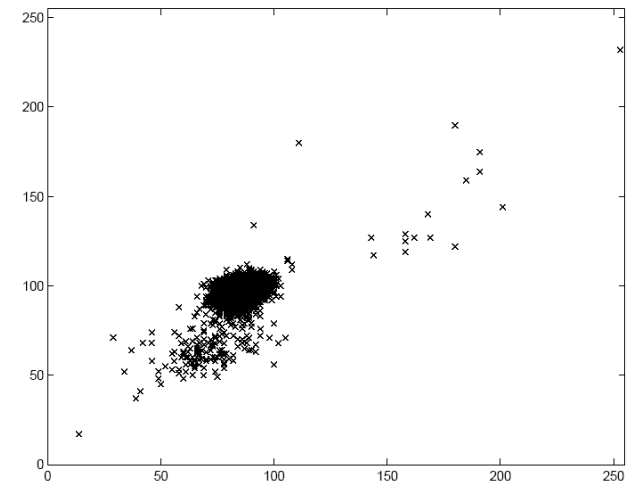
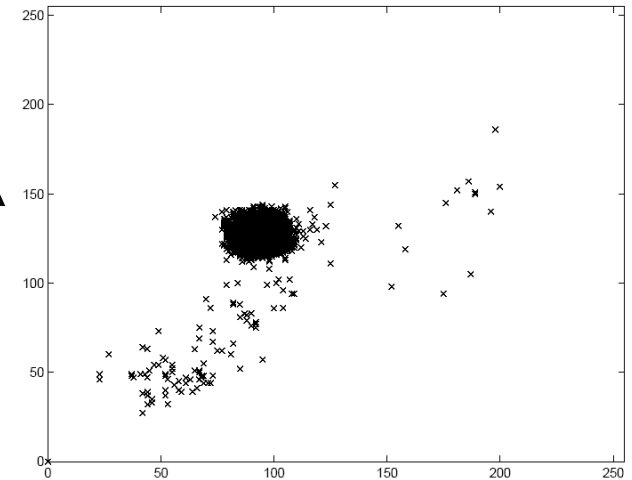
$$\sigma^2(x, y, t+1) = (1 - \alpha^2)\sigma^2(x, y, t) + \alpha^2(I(x, y, t+1) - \mu(x, y, t+1))^2$$

Mélanges de Gaussiennes - Introduction

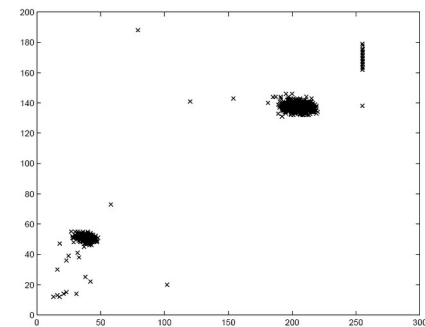
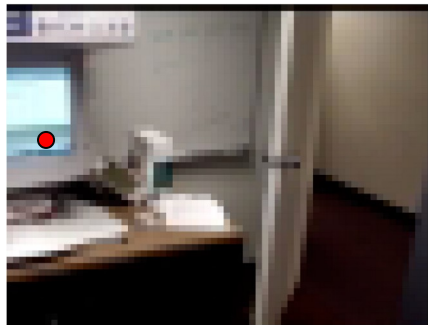
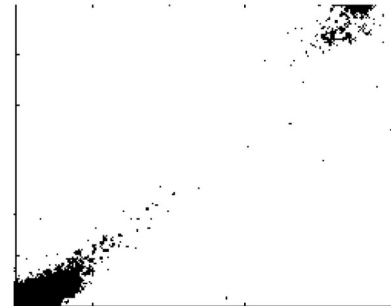
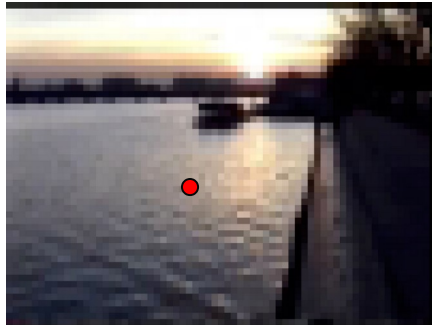
Même pixel

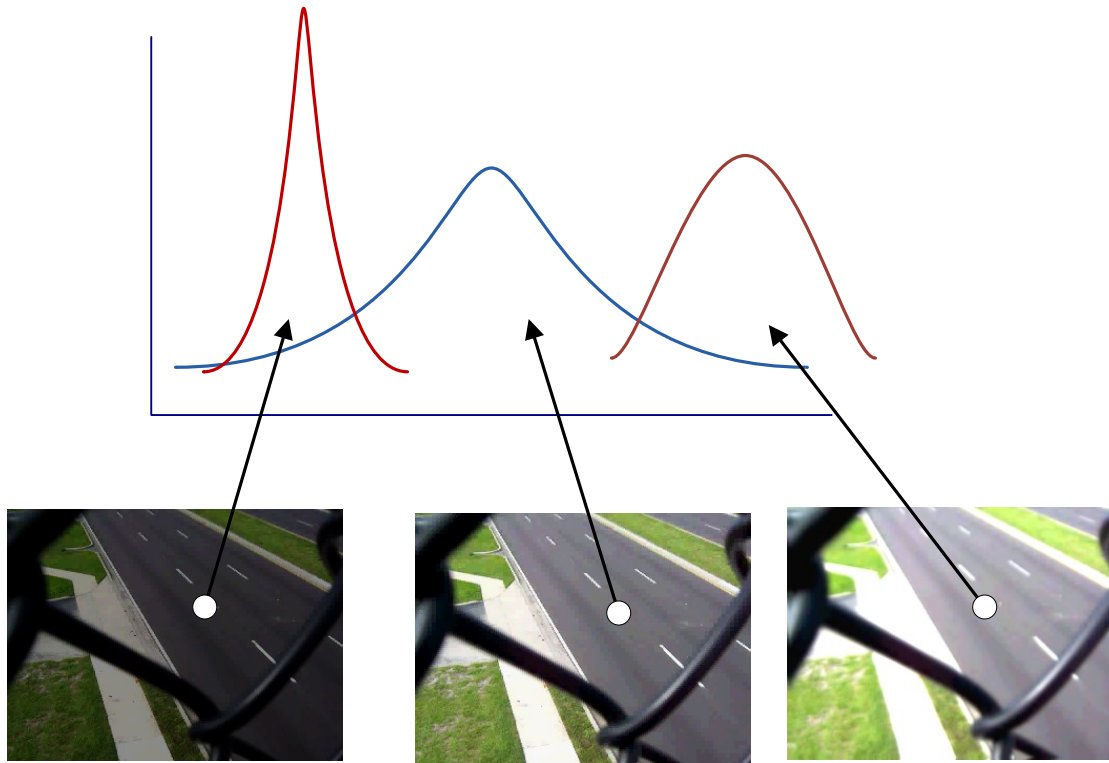


2 minutes plus tard

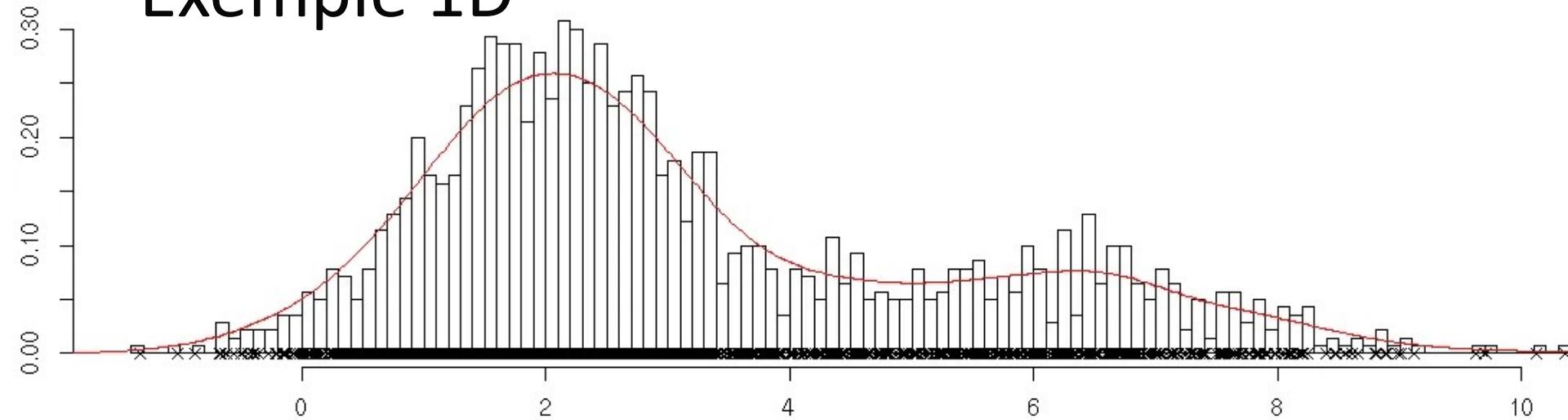


Mélange de Gaussiennes - Introduction

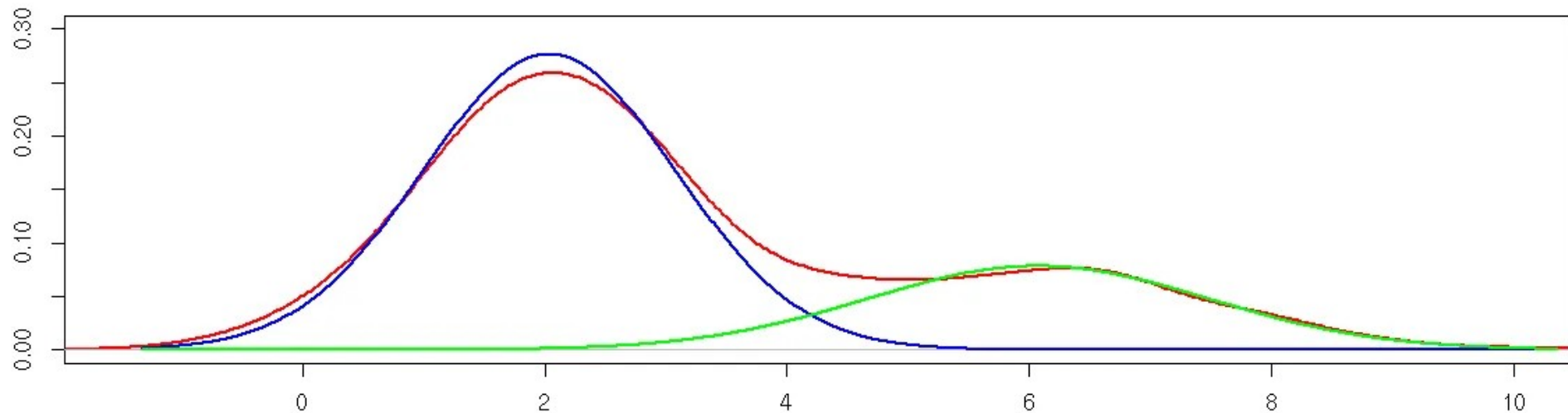




Exemple 1D



Plutôt que de chercher une loi compliquée on utilise une somme de Gaussiennes (ici 2 mais il peut y en avoir plus : K)



Mélange de Gaussiennes - Principe

Representation du Fond:

- Chaque pixel (i,j) à l'instant t consiste en K processus modélisés par une Gaussienne.

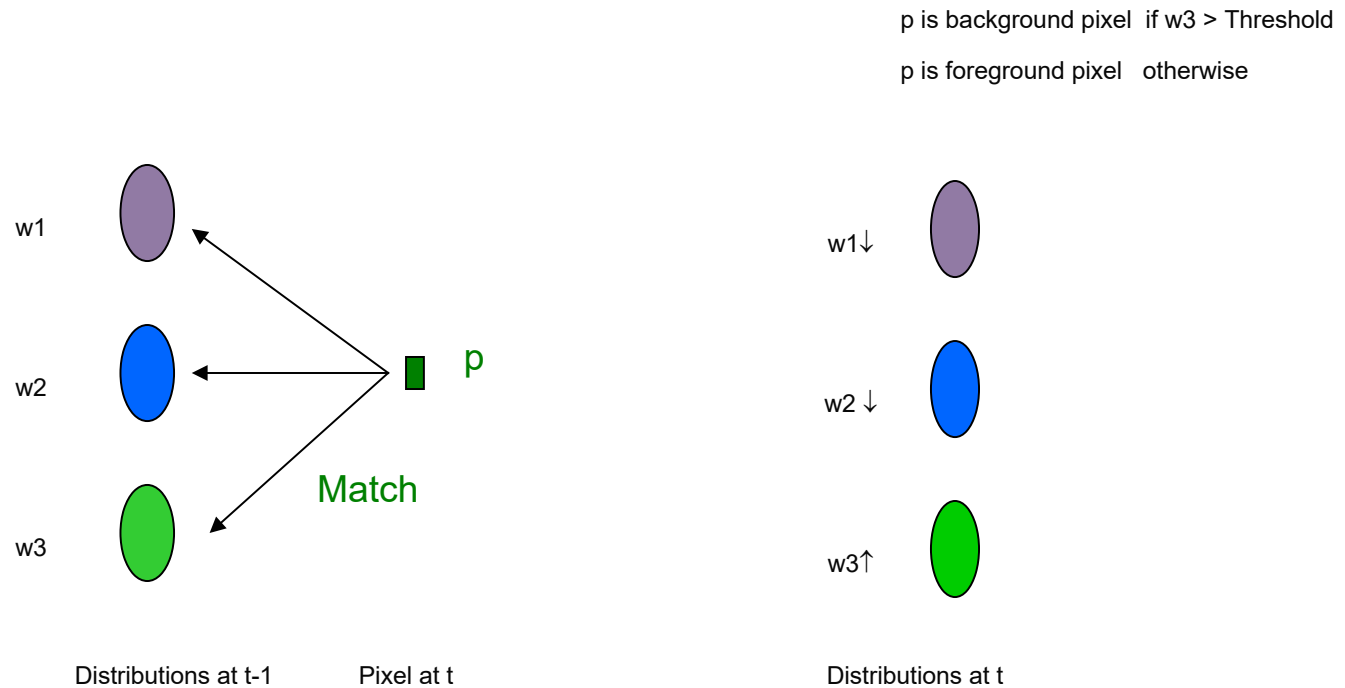
Algorithme :

- Initialisation des paramètres des K Gaussiennes avec un algorithme « EM » à partir d'une séquence d'apprentissage.
- A chaque image courante, pour chaque pixel, on calcule la distance entre ce pixel et chaque K Gaussiennes associées à ce pixel.

Mélange de Gaussiennes - Principe

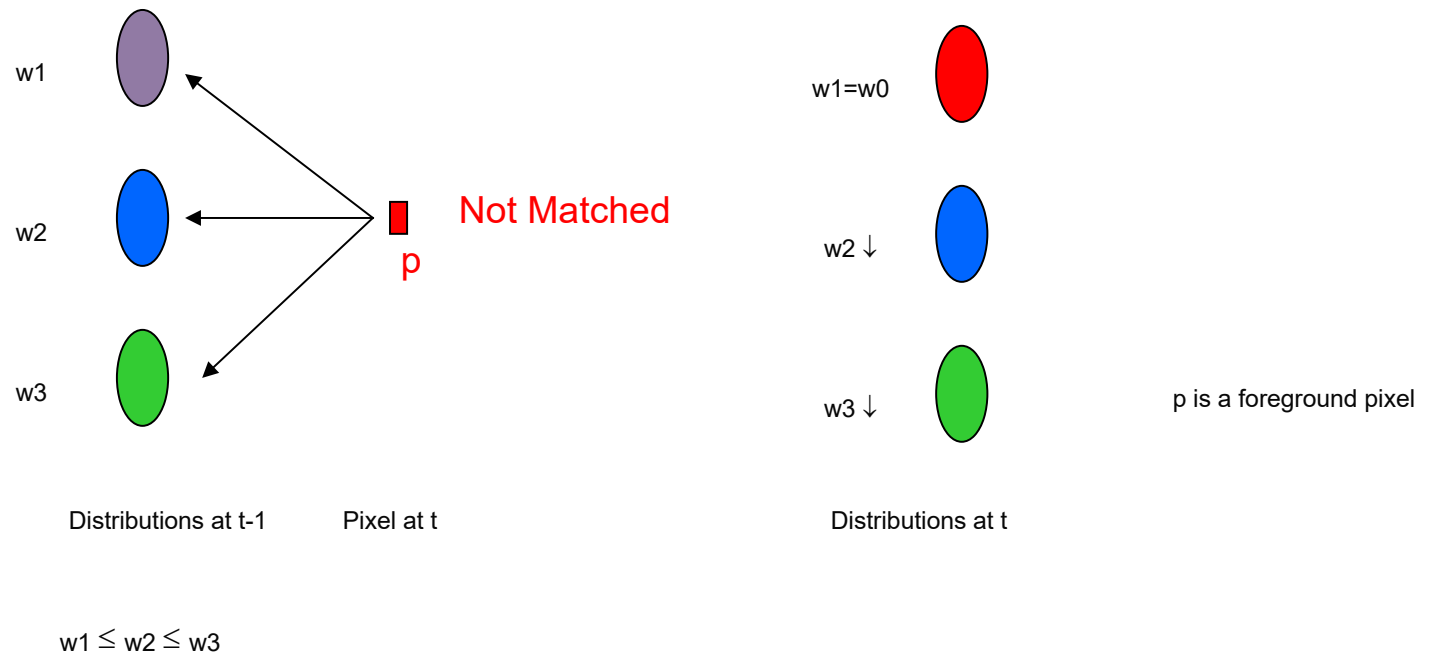
Cas 1 : Une correspondance est trouvée

Les poids sont ordonnés (croissants)



Mélange de Gaussiennes - Principe

Cas 2 : Aucune correspondance n'est trouvée



Mélange de Gaussiennes - Principe

Representation du fond

- Mixture of Gaussians (MOG) - Gaussians Mixture Model (GMM) :

$$\Pr(x_t) = \sum_{i=1}^K \omega_{i,t} \cdot N(x_t; \mu_{i,t}, \Sigma_{i,t})$$

K Nombre de gaussiennes

$\omega_{i,t}$ Estimation du poids de la i^{th} Gaussienne dans le mélange à l'instant t

$N(x_t; \mu_{i,t}, \Sigma_{i,t})$ Densité gaussienne

Mélange de Gaussiennes

Densité gaussienne :

$$N(x_t; \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{i,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - \mu_{i,t})^T (\Sigma_{i,t})^{-1} (x_t - \mu_{i,t})}$$

x_t Vecteur 3x1 valeur RGB du pixel

$\mu_{i,t}$ Vecteur 3x1 moyenne de la i^{eme} Gaussienne à l'instant t

$\Sigma_{i,t}$ Matrice de covariance 3x3 de la i^{eme} Gaussienne à t

Mélange de Gaussiennes

Pour simplifier, la matrice est diagonale :

$$\Sigma_{k,t} = \sigma_k^2 I$$

Ceci correspond à l'hypothèse (fausse ;)) d'indépendance entre les 3 canaux (i.i.d.) .

Mélange de Gaussiennes

Choix des paramètres

K est le nombre de modalité du fond : entre 3 to 5.

K souvent choisir en fonction des contraintes de mémoire.

Mélange de Gaussiennes

Test d'appartenance d'un pixel à une gaussienne

Si la distance de Mahalanobis est inférieure à un seuil :

$$\left((x_t - \mu_j)^T \cdot \Sigma_j^{-1} \cdot (x_t - \mu_j) \right) < (2.5 \sigma_j)^2$$

Ce seuil n'est pas très sensible et efficace:

- Il s'adapte à chaque gaussienne.
- Utile quand des régions de l'image ont des éclairages différents : les objets dans les zones à l'ombre sont en général moins bruités que dans les zones fortement éclairées (donc les variances sont plus faibles dans le premier cas).
- Si on choisissait un seuil uniforme les objets disparaîtraient en entrant dans les zones ombragées.

Mélange de Gaussiennes

Mise à jour du fond

Pour les gaussiennes qui matchent, la maj est :

$$\mu_{i,t} = (1 - \rho)\mu_{i,t-1} + \rho \cdot x_t$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(x_t - \mu_{i,t})^T \cdot (x_t - \mu_{i,t})$$

$$\omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha$$

$$\text{avec } \rho = \alpha \mathcal{N}(x_t | \mu_k, \sigma_k)$$

Pour les autres, moyennes et variances sont inchangées, seul le poids est diminué :

$$\omega_{j,t} = (1 - \alpha)\omega_{j,t-1}$$

Mélange de Gaussiennes

Si aucune gaussienne n'est matchée, la moins probable (poids le plus faible) est remplacée par :

$$\mu_{i,t} = x_t$$

$$\sigma_{i,t}^2 = \text{Large variance}$$

$$\omega_{i,t} = \text{Poids faible}$$

Mélange de Gaussiennes

Modélisation du fond

Si les K gaussienne sont ordonnées (croissant) selon ω_i/σ_i :

Poids maximaux et variances les plus faibles.

les B Gaussiennes qui dépassent un seuil T sont considérées comme le fond :

$$\operatorname{argmin}_b \left(\sum_{i=1}^b \omega_{i,t} > T \right)$$

T est une mesure de la proportion minimale qui est attribuée au fond.

Mélange de Gaussiennes

Permet de retenir les « meilleures » distributions

- Si T est petit on n'utilise, le plus souvent, une seule gaussienne. Permet d'accélérer le calcul mais peu réaliste en général.
- Si T est plus grand, une distribution multimodale causée par un mouvement répétitif du fond peut être pris en compte (ex: feuilles dans des arbres, drapeau au vent, feux clignotants...). Plusieurs couleurs sont donc incluses dans le modèle de fond.

Mélange de Gaussiennes

Significations des Parametres

Signification de K : Nombre de modes du fond.

Signification des poids :

- Proportion des données associées à chaque gaussienne
- Pertinence de la gaussienne dans les image récentes.

Mélange de Gaussiennes

Avantages :

- Prise en compte des changement lent d'illumination.
- Peut prendre en compte les changements du fond au bout d'un certain temps.
- Si l'initialisation inclus des objets en mouvement, le modèle va se corriger avec le temps.

Mélange de Gaussiennes

Inconvénients :

- Difficultés avec les changements rapides.
- Si l'initialisation est mauvaise, il faut du temps pour corriger.
- Les ombres restent difficiles à isoler.

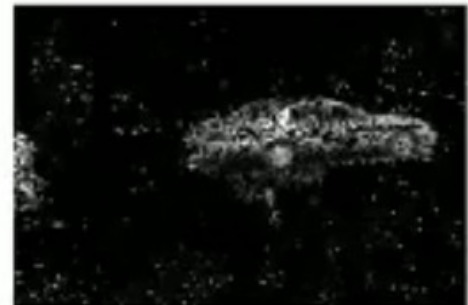
Orig [441/500]



Background



Foreground



<https://www.youtube.com/watch?v=qiZ1g65UiCw>

Modèles Statistiques – Première Génération

- Gaussienne [Wren 1997]
- Mélange de Gaussiennes [Stauffer 1999]
- Estimation de Densité de Noyau [Elgammal 2000]
- **Eigenbackground [Oliver 2000]**

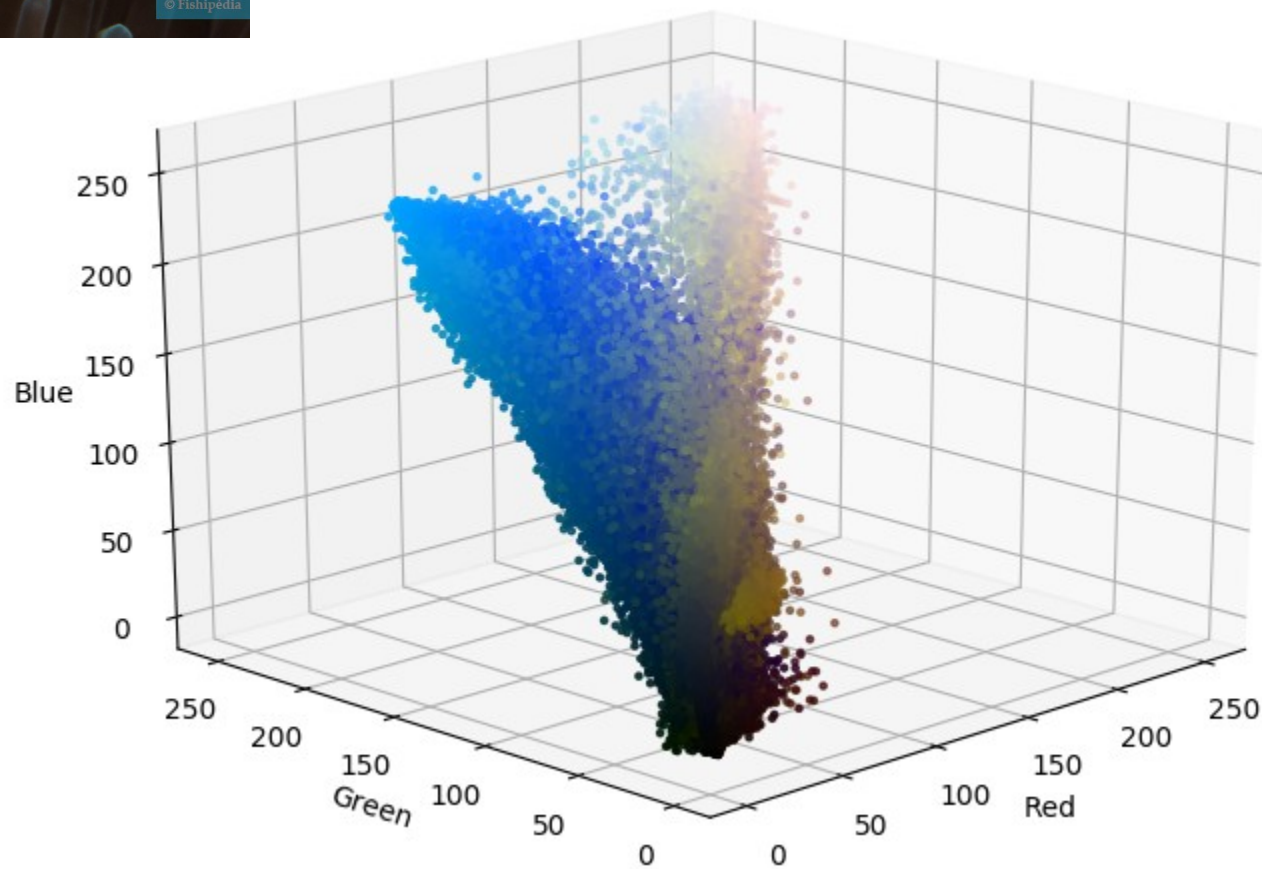
Eigenbackgrounds – Principe

Modélisation du fond à partir des valeurs propres:
Oliver et al [Oliver 2000].

Representation du fond

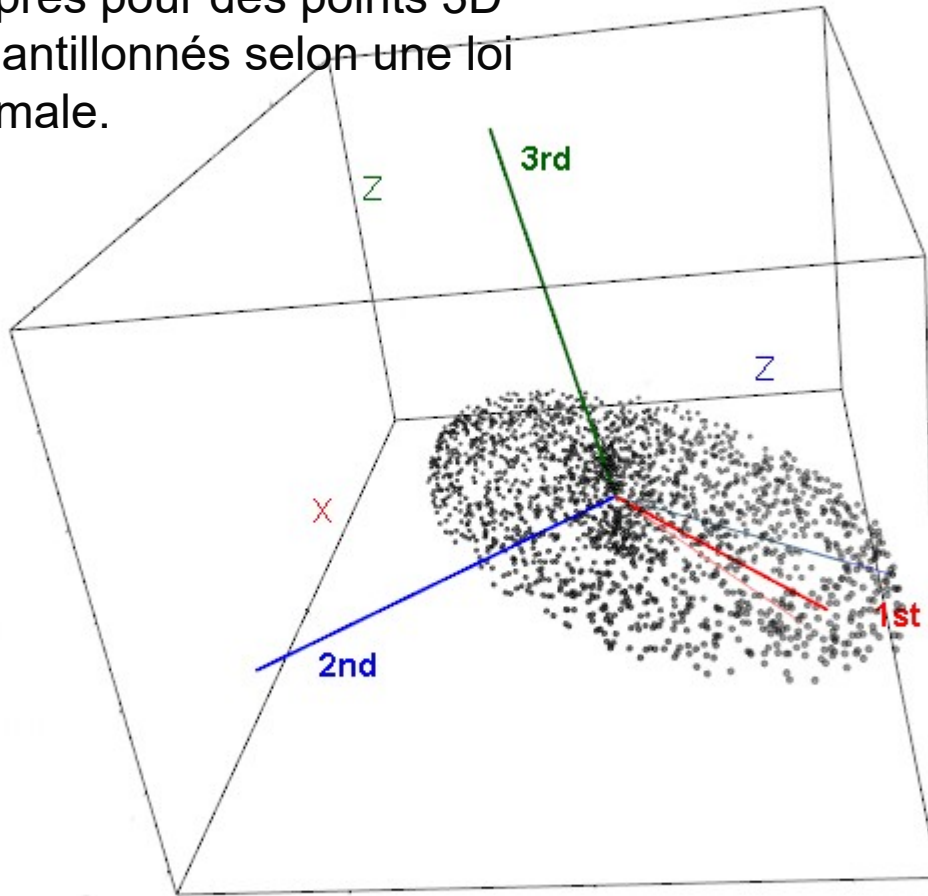
On applique une Analyse en Composantes Principale (ACP) sur un jeu d'images d'apprentissage et le fond est représenté par l'image moyenne et p valeurs propres significatives : les **eigenbackgrounds**.

Quand les eigenbackgrounds sont calculés, une image en entrée peut être projetée dans l'espace des eigenbackgrounds, ce qui permet de reconstruire une image de référence.

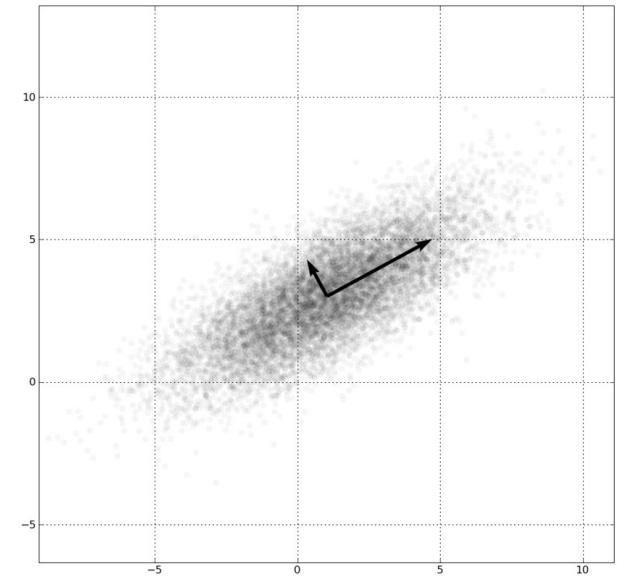


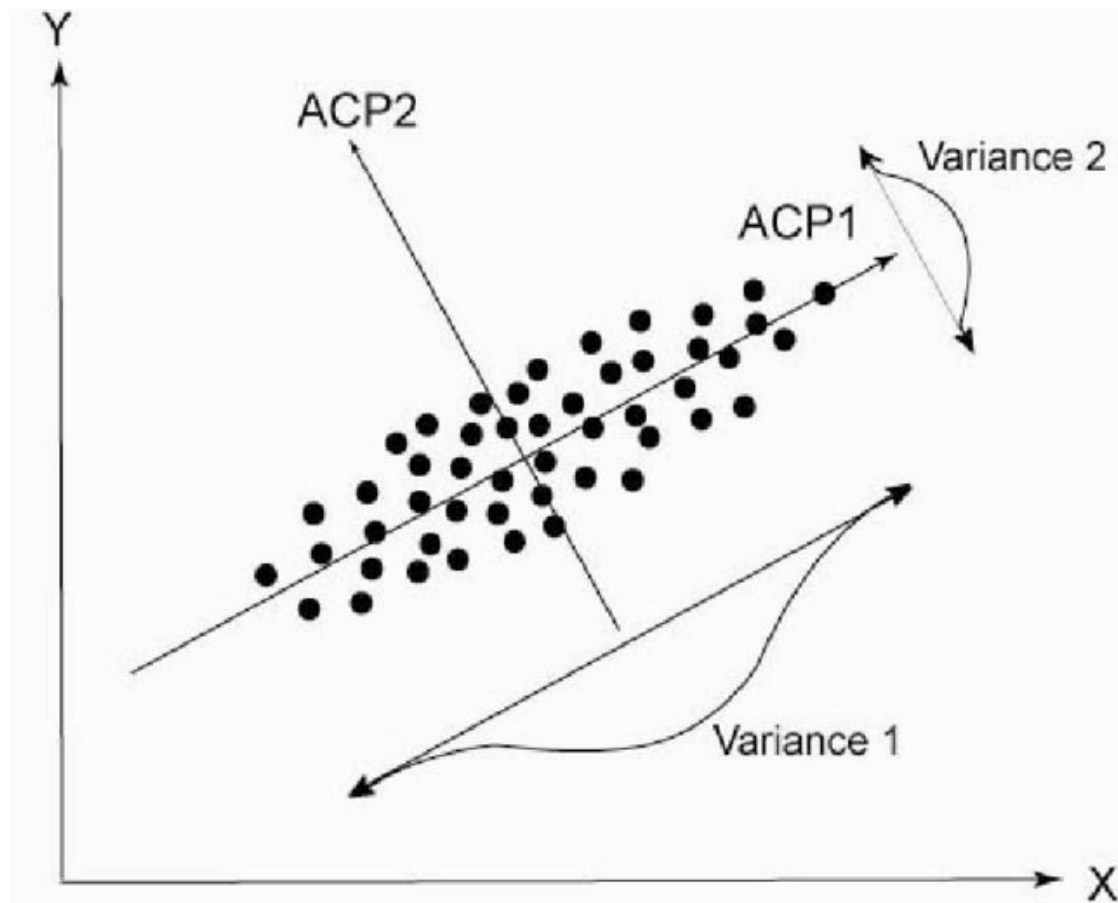
Analyse en Composantes Principales (ACP)

Vecteurs propres et valeurs propres pour des points 3D échantillonnés selon une loi normale.



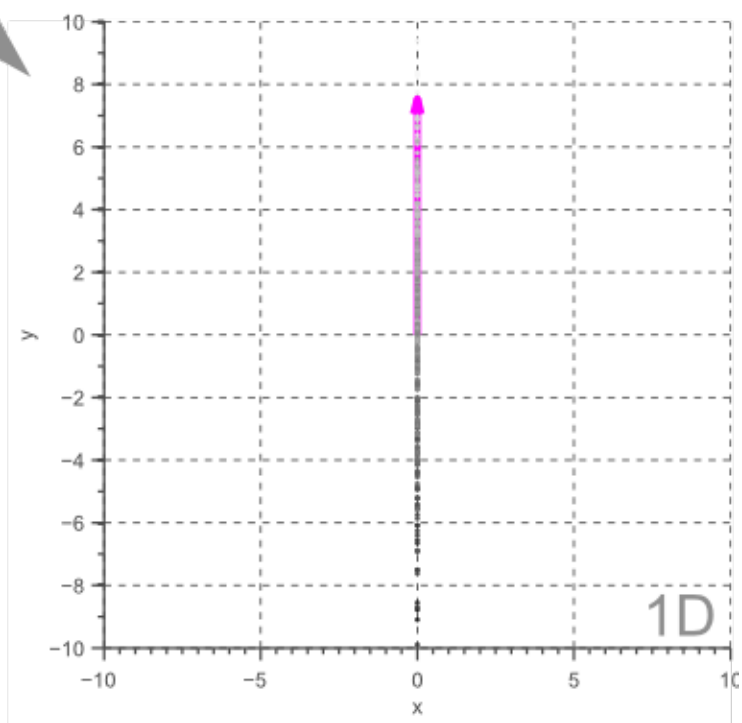
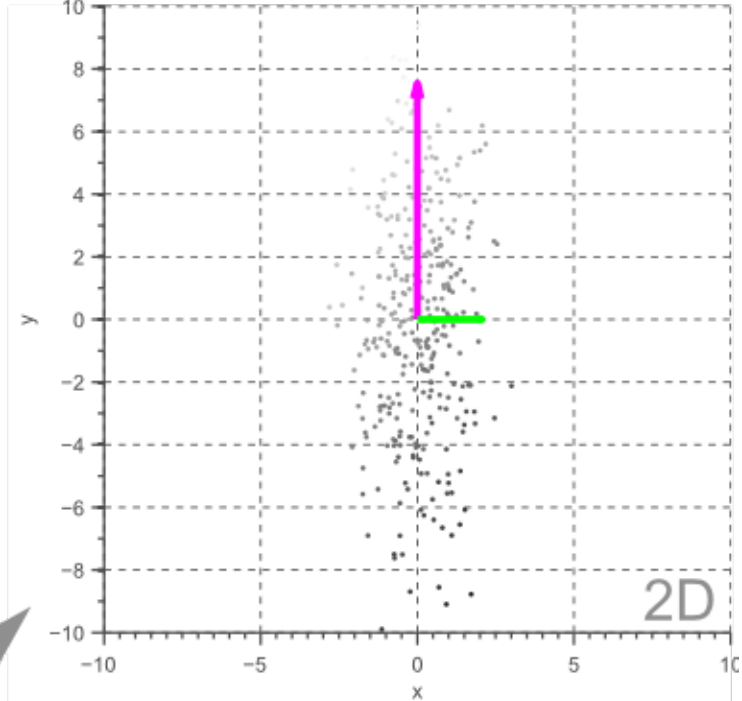
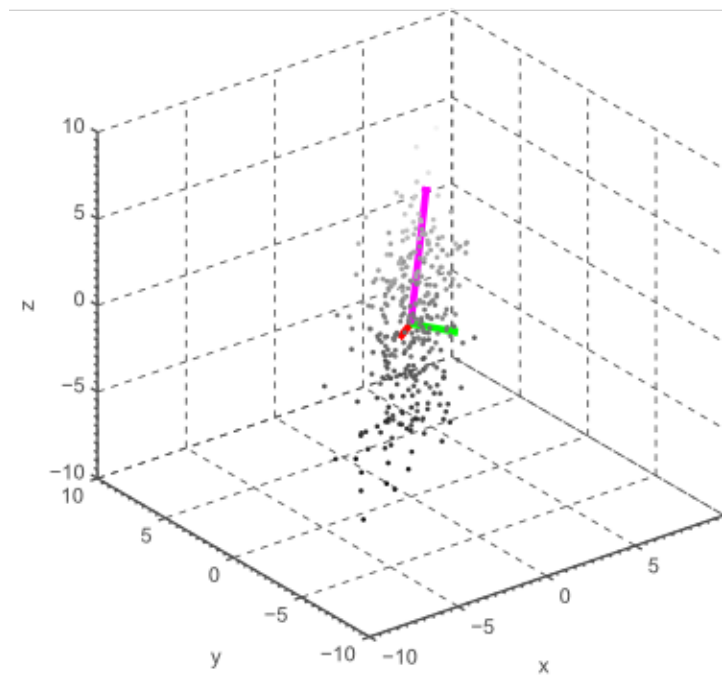
En 2D

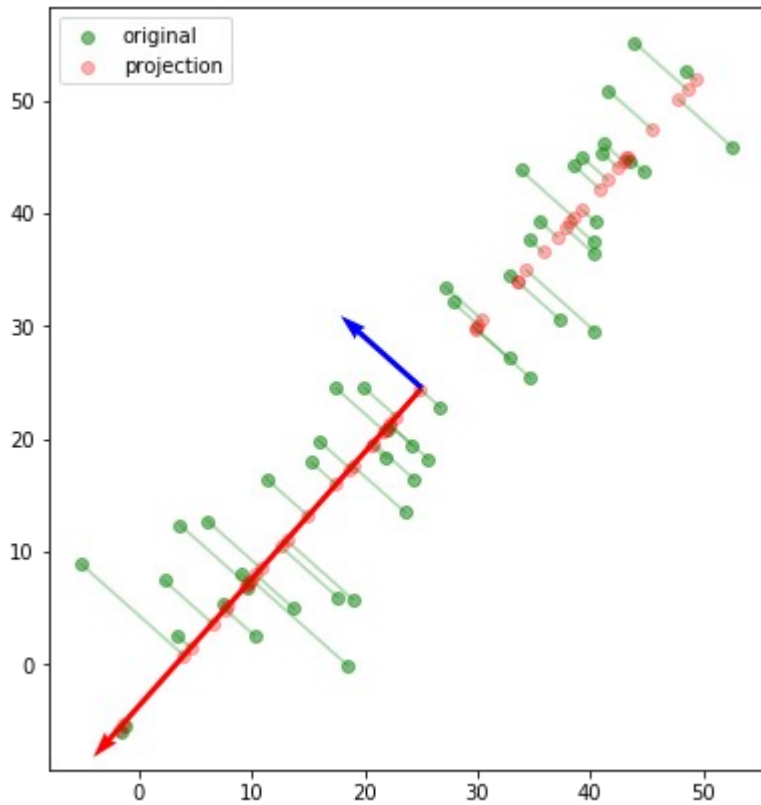




« L'ACP va déterminer les deux axes qui expliquent le mieux la dispersion de l'objet, interprété comme un nuage de points. Elle va aussi les ordonner par inertie expliquée, le second axe étant perpendiculaire au premier. » Wikipedia

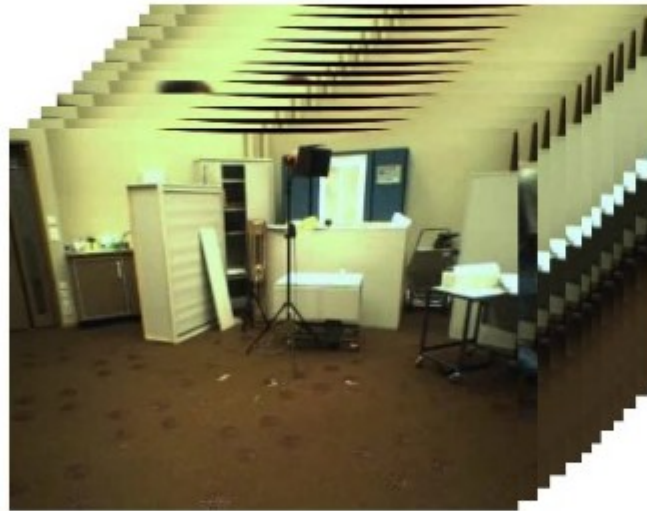
Idem en dimensions >2







(a) Une séquence de 200 trames vue comme une succession d'images 2D. (Seules les trames 1, 27, 93, 151 et 200 sont représentées.)



(b) Vue compacte de la même séquence obtenue en empilant toutes les images le long de la troisième dimension.

FIG. 4.1 – Exemple de séquence vidéo avec sa représentation en 2D (a) et volumique (b).

Eigenbackgrounds – Principe

Classification

Les pixels d'avant plan (foreground) sont obtenus en calculant la différence entre l'image donnée en entrée et l'image reconstruite avec un nombre limité de vecteurs propres.

Eigenbackgrounds

Entraînement

Les n images d'entraînement sont placées (dépliées ») sous forme de vecteurs colonnes $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_n$, et le vecteur (image) moyen est $\boldsymbol{\varphi}$. La différence entre une image et la moyenne est notée $\boldsymbol{\psi}_i = \mathbf{y}_i - \boldsymbol{\varphi}$.

L'ACP est appliquée sur ces vecteurs, elle permet d'obtenir p *vecteurs propres* orthonormés (eigenvectors) $\boldsymbol{\mu}_k$ et autant de valeurs propres (eigenvalues) λ_k à partir de la matrice de covariance :

$$C = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T = \mathbf{A} \mathbf{A}^T$$

où $\mathbf{A} = [\boldsymbol{\psi}_1 \boldsymbol{\psi}_2 \boldsymbol{\psi}_3 \dots \boldsymbol{\psi}_n]$ (taille $m \times n$) , et $\mathbf{A} \mathbf{A}^T \boldsymbol{\mu} = \lambda \boldsymbol{\mu}$ (*vecteurs propres*)

Remarque :

les $\mathbf{y}_i$ sont des vecteurs de dimension = le nombre de pixels de l'image (ex : $1000 \times 1000 = 10^6$ donc C est de dimension $\mathbf{m} = 10^6 \times 10^6$)

Eigenbackgrounds

La matrice **C étant grande**, on utilise une astuce technique en calculant valeurs et vecteurs propre sur **C^T**.

C^T est de taille $n \times n \ll m \times m$

C^T = A^TA, a pour vecteurs propres v_k et pour valeurs propres ξ_k . On se ramène à μ_k et λ_k par :

$$\mu_k = \frac{Av_k}{\sqrt{\xi_k}} \quad \lambda_k = \xi_k$$

Demo :

A^TAv = ξv par définition (v.p.)

AA^T(Av) = ξ(Av) par identification à **AA^Tμ = λ μ** :

- $\xi = \lambda$ et
- **Av = μ** mais non normalisés.

Eigenbackgrounds

Phase de test

L'image courante (γ_i) contenant les objets de premier plan peut être approximée par le fond moyen (φ) et la somme pondérée des eigenbackgrounds μ_k .

On utilise pour cela la matrice des eigenbackgrounds

$\Delta = [\mu_1, \mu_2, \mu_3, \dots, \mu_p]$, et les poids dans l'espace des eigenbackgrounds de l'image γ_i est calculée par :

$$w = (\gamma_i - \varphi)^T \Delta$$

Eigenbackgrounds

$\mathbf{w}$ est rétroprojeté dans l'espace image pour créer une image de fond de référence :

$$\gamma_i^R = \Delta \mathbf{w}^T + \varphi$$

Comme la matrice des eigenbackground γ_i^R représente le fond (et donc pas les détails) les petits objets en mouvement n'y figurent pas.

Eigenbackgrounds

Detection de l'avant plan (foreground)

Il suffit de calculer la différence entre l'image courante et l'image de fond reconstruite pour obtenir les objets en mouvement et de seuiller :

$$\left| \gamma_i - \gamma_i^R \right| > T$$

γ_i Image courante

γ_i^R Image reconstruite

<https://www.youtube.com/watch?v=4c1ILY3ta08>



DPEigenbackgroundBGS

Frame n. 1179

Bibliographie

- [Wren 1997] - C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pffinder : Real-Time Tracking of the Human Body", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, No. 7, pages 780 –785 , July 1997.
- [Stauffer 1999] - C. Stauffer, E. Grimson, "Adaptive background mixture models for real-time tracking", Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1999), pages 246-252, 1999.
- [Elgammal 2000] - A. Elgammal, D. Harwood, L. Davis, "Non-parametric Model for Background Subtraction", 6th European Conference on Computer Vision (ECCV 2000), Dublin, Ireland, June 2000.
- [Oliver 2000]- N. Oliver, B. Rosario, A. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions", PAMI Special Issue on Visual Surveillance and Monitoring, PAMI 2000, August 2000.