

AIV2 – Description de la vidéo

Pierre Tirilly

Master Informatique, parcours RVA – Université de Lille

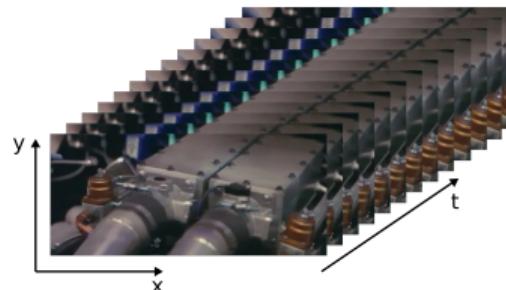
Dans les épisodes précédents de AIV2

Reconnaissance des images statiques : Descripteurs d'apparence + classifieur

- ▶ Descripteurs globaux (ex : histogrammes de couleurs)
- ▶ Descripteurs locaux :
 - ▶ Détection de points d'intérêt (ex. : Harris, DoG)
 - ▶ Description des régions attenantes (ex. : HOG/SIFT)
 - ▶ Agrégation de descripteurs locaux (ex. : BOVW, VLAD)
- ▶ Structuration géométrique des descripteurs (ex. : découpage en grille)

Vidéo : Séquence d'images statiques ("Cube" vidéo)

- ▶ Deux dimensions spatiales + une dimension temporelle
- Nouvelle information visuelle : mouvement (cf. flux optique)



Objectifs du jour

Objectif : Reconnaissance (d'actions, d'expression faciales, de personnes, etc.) dans des vidéos

Questions :

1. Comment détecter les régions en mouvement ?
2. Comment décrire le mouvement dans la vidéo ?

1. Détection des régions en mouvement

Points d'intérêt spatio-temporels (*Space-time interest points – STIP*)

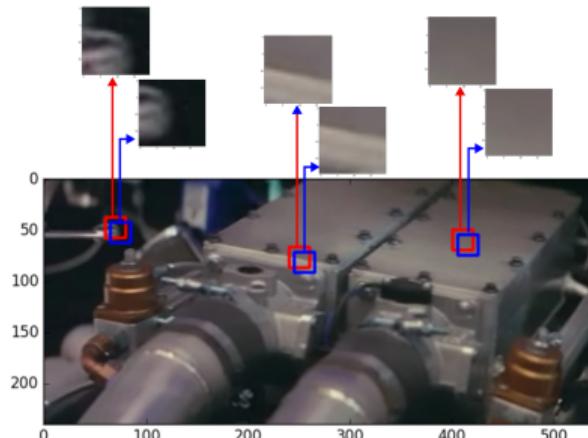
Problématique : Identifier des régions singulières dans la vidéo

- ▶ Au sens de l'apparence
- ▶ Au sens du mouvement



Apparence singulière : détecteur de Harris

Région singulière ? Définie comme une région variante sous les translations



Formalisation :

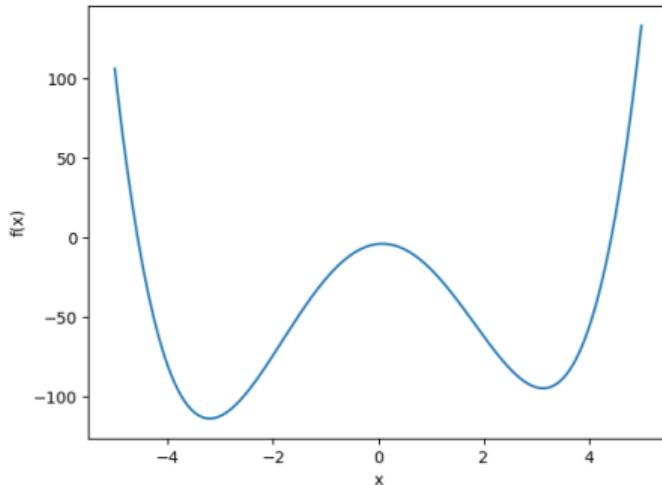
$$R(I, \mathcal{R}, d_x, d_y) = \sum_{(x,y) \in \mathcal{R}} (I(x + d_x, y + d_y) - I(x, y))^2$$

- ▶ $R(I, \mathcal{R}, d_x, d_y)$ indique la variabilité de la région \mathcal{R} de I sous la translation (d_x, d_y)
- ▶ On cherche les régions \mathcal{R} qui maximisent $R(I, \mathcal{R}, d_x, d_y)$ pour tout (d_x, d_y)

Détecteur de Harris : développement

Objectif : Trouver les régions \mathcal{R} maximisant $R(I, \mathcal{R}, d_x, d_y)$

- ▶ Pas de formule analytique pour R
 - ▶ Impossible à évaluer pour tout (d_x, d_y)
- Approximation linéaire (Taylor, ordre 1)



Reformulation :

$$R(I, \mathcal{R}, d_x, d_y) = \sum_{(x,y) \in \mathcal{R}} (I(x + d_x, y + d_y) - I(x, y))^2$$

$$R(I, \mathcal{R}, d_x, d_y) = \sum_{(x,y) \in \mathcal{R}} (I(x, y) + d_x \cdot \frac{\partial I}{\partial x}(x, y) + d_y \cdot \frac{\partial I}{\partial y}(x, y) + r(x, y) - I(x, y))^2$$

$$R(I, \mathcal{R}, d_x, d_y) = \sum_{(x,y) \in \mathcal{R}} (d_x \cdot \frac{\partial I}{\partial x}(x, y) + d_y \cdot \frac{\partial I}{\partial y}(x, y))^2$$

Détecteur de Harris : développement

Développement de $R(I, \mathcal{R}, d_x, d_y)$:

$$R(I, \mathcal{R}, d_x, d_y) = \sum_{(x,y) \in \mathcal{R}} (d_x \cdot \frac{\partial I}{\partial x}(x, y) + d_y \cdot \frac{\partial I}{\partial y}(x, y))^2$$

$$R(I, \mathcal{R}, d_x, d_y) = \sum_{(x,y) \in \mathcal{R}} (d_x^2 \cdot \frac{\partial I^2}{\partial x}(x, y) + 2d_x d_y \cdot \frac{\partial I}{\partial x}(x, y) \cdot \frac{\partial I}{\partial y}(x, y) + d_y^2 \cdot \frac{\partial I^2}{\partial y}(x, y))$$

$$R(I, \mathcal{R}, d_x, d_y) = [d_x \ d_y] \cdot \begin{bmatrix} \sum \frac{\partial I^2}{\partial x} & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I^2}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} d_x \\ d_y \end{bmatrix}$$

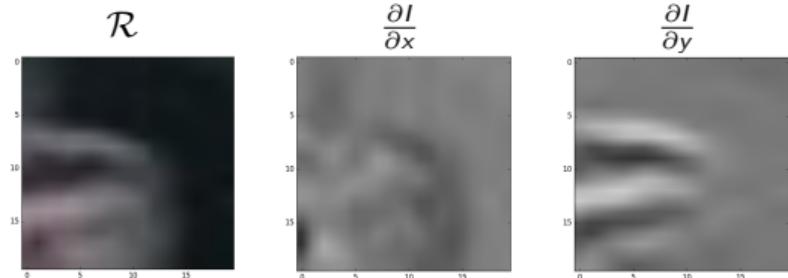
→ $R(I, \mathcal{R}, d_x, d_y)$ dépend du tenseur de structure :

$$S = \begin{bmatrix} \sum \frac{\partial I^2}{\partial x} & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I^2}{\partial y} \end{bmatrix}$$

Distribution des gradients

Tenseur de structure : Décrit la distribution du gradient ∇I dans \mathcal{R}

$$S = \begin{bmatrix} \sum \frac{\partial I}{\partial x}^2 & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I}{\partial y}^2 \end{bmatrix}$$



Valeurs propres de S : Échelle de la distribution de ∇I dans ses directions principales

$$S = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Identification des types de régions

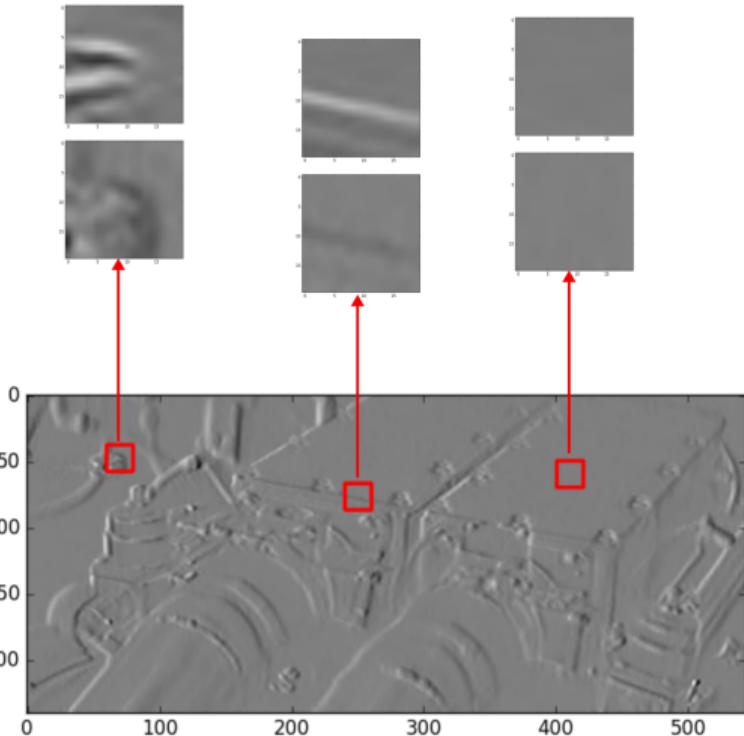
Tenseur de structure :

$$S = \begin{bmatrix} \sum \frac{\partial I}{\partial x}^2 & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I}{\partial y}^2 \end{bmatrix}$$

$$S = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Cas possibles :

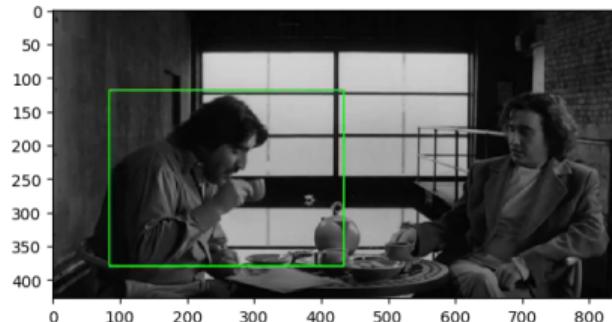
- ▶ Région uniforme : $\lambda_1 \approx 0, \lambda_2 \approx 0$
- ▶ Arête : $\lambda_1 \nearrow, \lambda_2 \approx 0$
- ▶ Contours complexes (coins, etc.) :
 $\lambda_1 \nearrow, \lambda_2 \nearrow$



Réponse du détecteur

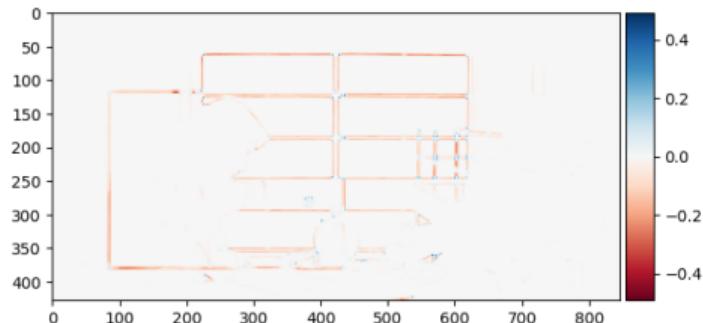
Mesure de singularité :

$$\begin{aligned} R'(I, \mathcal{R}) &= \det(S) - k \cdot \text{trace}(S)^2 \\ R'(I, \mathcal{R}) &= \lambda_1 \cdot \lambda_2 - k \cdot (\lambda_1 + \lambda_2)^2 \end{aligned}$$

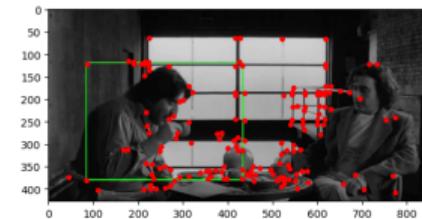
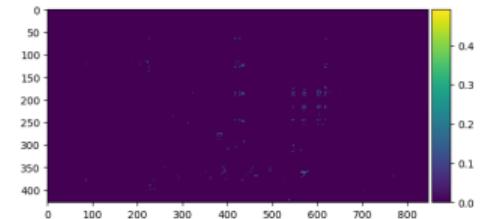
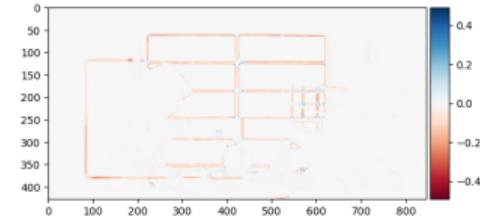
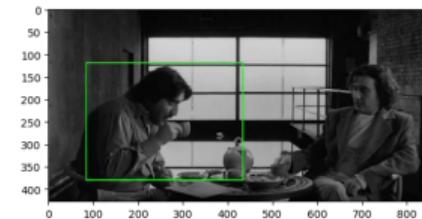


Cas possibles :

- ▶ Région uniforme : $R'(I, \mathcal{R}) \approx 0$
- ▶ Arête : $R'(I, \mathcal{R}) < 0$
- ▶ Contours complexes (coins, etc.) : $R'(I, \mathcal{R}) > 0$



Calcul des régions d'intérêt



Étapes :

1. Calcul de $R'(I, \mathcal{R})$ pour tout \mathcal{R}
2. Conservation des \mathcal{R} t.q. $R'(I, \mathcal{R}) > \theta$
3. Suppression des non-maxima locaux

Paramètres :

- ▶ $k \in [0.04, 0.06]$
- ▶ θ

STIP : Space-time interest points

Principe : Une région est singulière si elle varie sous une translation *spatio-temporelle*

Formalisation :

$$R(I, \mathcal{R}, d_x, d_y, d_t) = \sum_{(x,y,t) \in \mathcal{R}} (I(x + d_x, y + d_y, t + d_t) - I(x, y, t))^2$$

Développement : Similaire au détecteur de Harris, en 3 dimensions (2D + t)

$$R(I, \mathcal{R}, d_x, d_y, d_t) = [d_x \ d_y \ d_t] \cdot \begin{bmatrix} \sum \frac{\partial I}{\partial x}^2 & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial t} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I}{\partial y}^2 & \sum \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial t} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial t} & \sum \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial t} & \sum \frac{\partial I}{\partial t}^2 \end{bmatrix} \cdot \begin{bmatrix} d_x \\ d_y \\ d_t \end{bmatrix}$$

Tenseur de structure :

$$S = \begin{bmatrix} \sum \frac{\partial I}{\partial x}^2 & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial t} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \sum \frac{\partial I}{\partial y}^2 & \sum \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial t} \\ \sum \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial t} & \sum \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial t} & \sum \frac{\partial I}{\partial t}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Fonction de singularité :

$$R'(I, \mathcal{R},) = \det(S) - k \cdot \text{trace}(S)^3$$

$$R'(I, \mathcal{R},) = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 - k \cdot (\lambda_1 + \lambda_2 + \lambda_3)^3$$

Conclusion sur les points d'intérêt spatio-temporels

STIP : Extension des points de Harris au domaine spatio-temporel

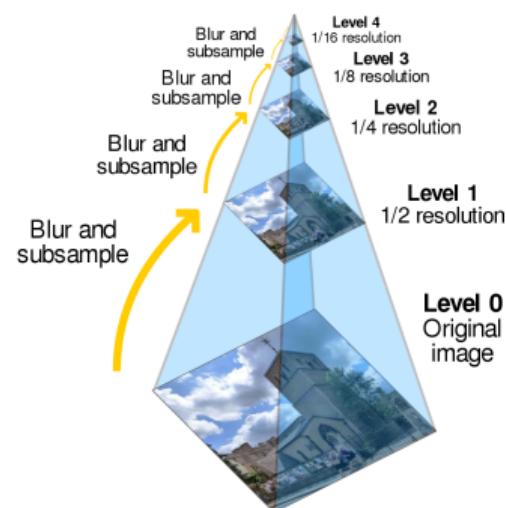
- ▶ Régions variables sous une translation
- ▶ Basé sur le tenseur de structure local
- ▶ Détection de "coins" en mouvement

Autres méthodes : Extension de méthodes spatiales au domaine spatio-temporel

- ▶ Filtres de Gabor ("Cuboïdes") [Dollár et al.]
- ▶ Hessian3D [Willems et al.]
- ▶ [Échantillonage dense : grille 3D dans le cube vidéo]
- ▶ ...

Détection multi-échelle : Approche *scale-space*

- ▶ Filtre gaussien + redimensionnement à différentes échelles
- ▶ Détection des points à chaque échelle



Source : [wikipedia.org](https://en.wikipedia.org)

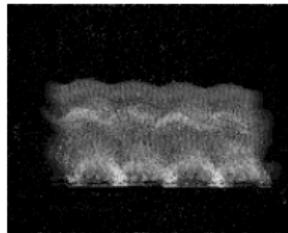
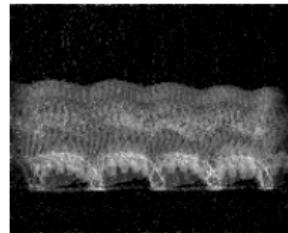
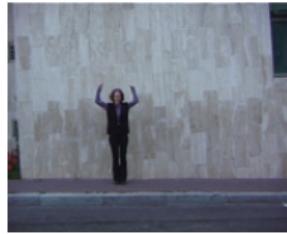
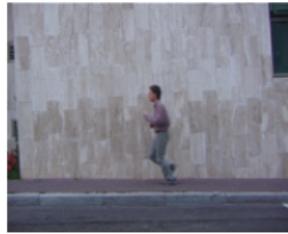
2. Description du mouvement

2.1. Description globale du mouvement

MHI : Motion History Images

Principe : Trame “résumée” du mouvement dans le cube vidéo

Descripteur : Calcul de moments statistiques sur l'historique (moyenne, variance, etc.)



Run

Run

Wave

MHI : Calcul

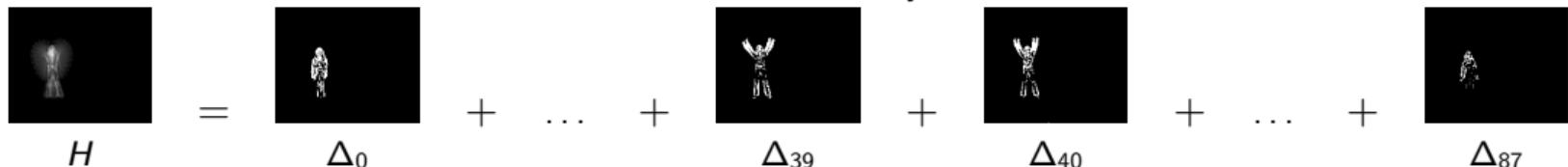
Trames vidéos:



Mouvement: Différence d'images seuillée $\Delta_t = \begin{cases} 1 & \text{si } I_{t+1} - I_t > \theta \\ 0 & \text{sinon} \end{cases}$



Historique de mouvement: Somme sur les trames $H = \sum_t \Delta_t$



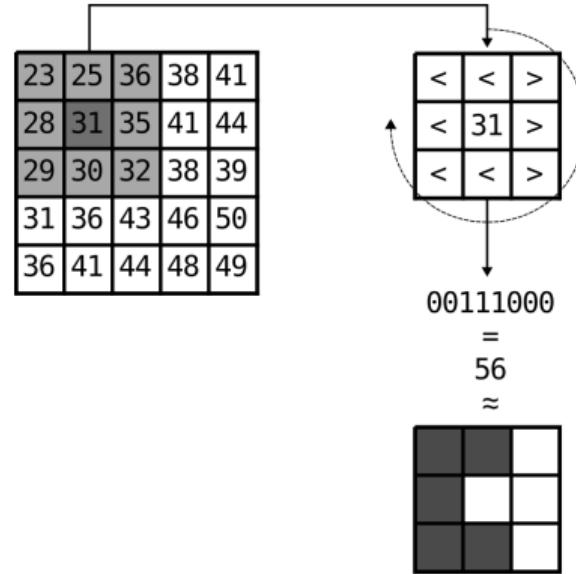
2.2. Description locale du mouvement

LBP spatio-temporels

Rappel : LBP (*Local Binary Patterns*)

- ▶ Signature binaire du voisinage d'un point
- ▶ Nombre de patterns : 2^n , n = nombre de points voisins = taille de la signature
- ▶ Calcul d'un histogramme des LBP extraits en chaque pixel

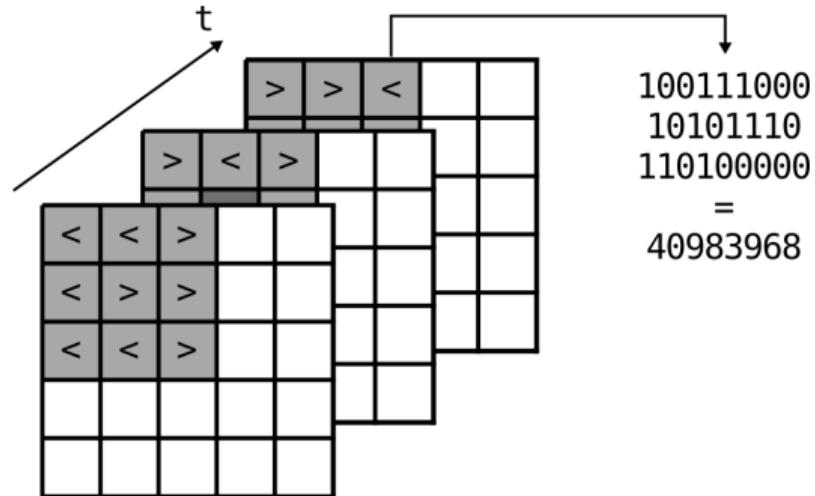
Problématique : Comment calculer ces signatures dans un voisinage 2D+ ?



LBP spatio-temporels : VLBP

VLBP : *Volume Local Binary Patterns* [Zhao & Pietikäinen, 2007]

- ▶ Voisinage en 3 dimensions
- ▶ Calcul de LBP sur ce voisinage
- ▶ Calcul d'un histogramme des LBP extraits en chaque pixel

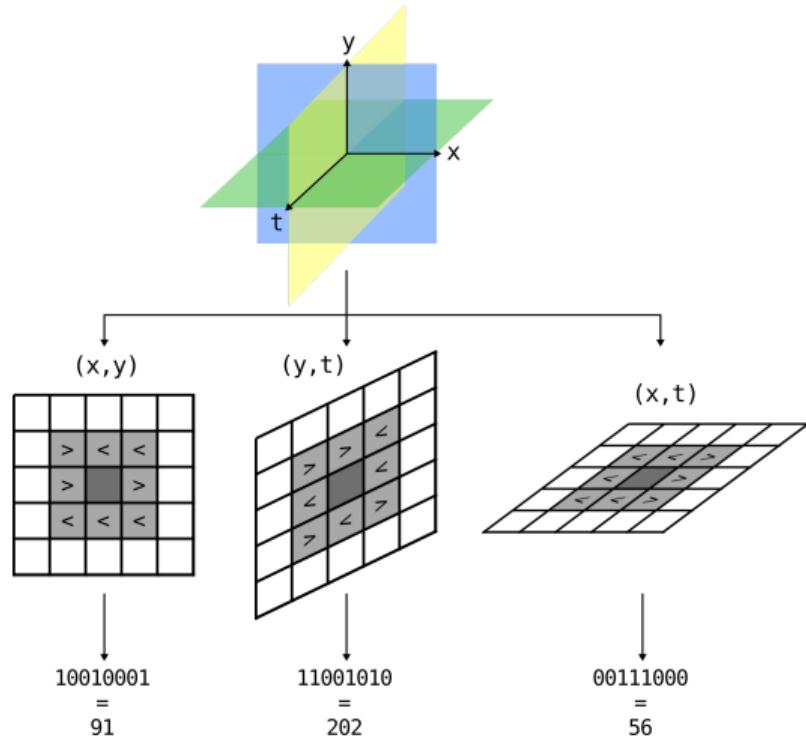


Problème : Nombre de patterns : 2^{3n+2} , $n = \text{nombre de points voisins sur un plan}$
→ Histogramme de dimension trop élevée

LBP spatio-temporels : LBP-TOP

LBP-TOP : *Local binary patterns on three orthogonal planes* [Zhao & Pietikäinen, 2007]

- ▶ Voisinage limité à trois plans : (x, y) , (x, t) , (y, t)
- ▶ Calcul d'une signature LBP sur chaque plan
- ▶ Concaténation des histogrammes de chaque plan
- ▶ Nombre de patterns : 3×2^n , $n =$ nombre de points voisins sur un plan



Histogrammes de gradients spatio-temporels

Rappel : Histogrammes de gradients

1. Calcul du gradient spatial $\nabla I = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}]^T$
2. Découpage de la région \mathcal{R} en $r_x \times r_y$ sous-régions
3. Pour chaque sous-région \mathcal{R}_i :

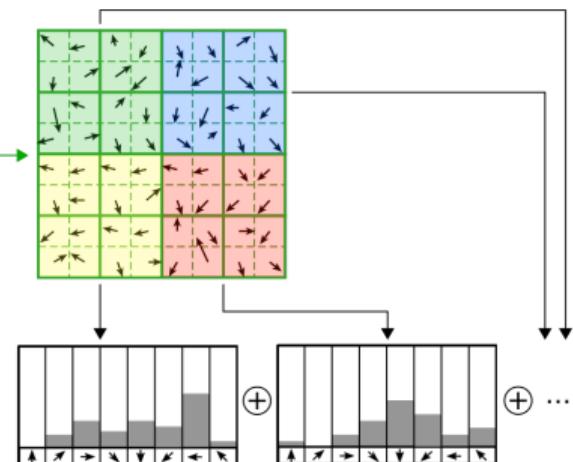
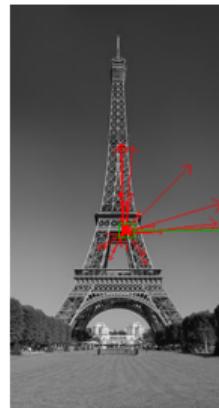
- (a) Quantifier les orientations des vecteurs de gradient $d = [d_x, d_y]^T$ en n_q bins :

$$q(d) = \left\lfloor \frac{\alpha}{n_q} \right\rfloor, \quad \alpha = \arctan \left(\frac{d_y}{d_x} \right)$$

- (b) Calculer l'histogramme pondéré :

$$h^{(i)}(b) = \sum_{d | q(d)=b} ||d||_2, \quad \forall b \in [0, n_q[$$

4. Concaténer les histogrammes des régions \mathcal{R}_i :
$$h = [h^{(1)}, h^{(2)}, \dots, h^{(r_x \times r_y)}]$$
5. Normaliser l'histogramme (L1, L2, etc.)



Problématique : Comment calculer un histogramme d'orientations dans un volume 2D+t ?

Histogrammes de gradients spatio-temporels

Option 1 : Cumul des histogrammes spatiaux [Laptev et al., 2008]

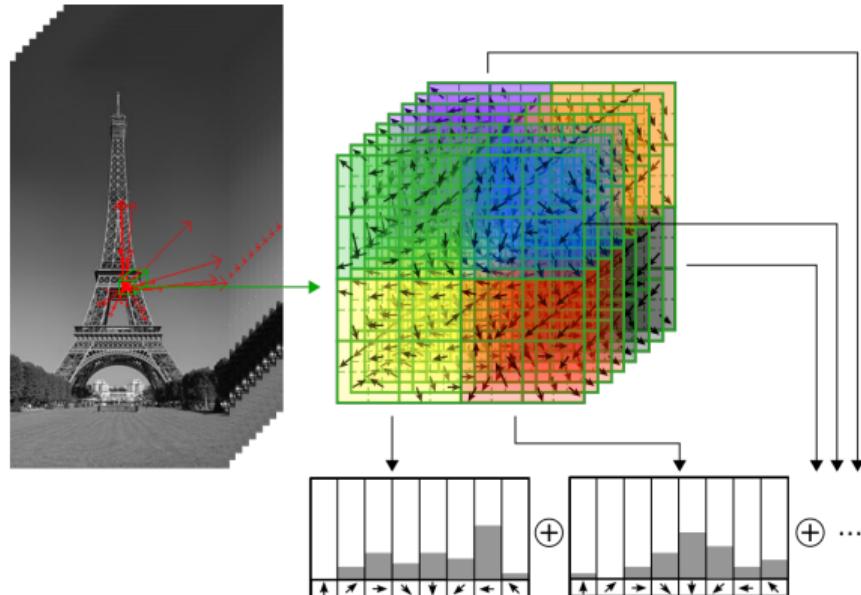
1. Calcul du gradient spatial $\nabla I = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}]^T$
2. Découpage de la région \mathcal{R} en $r_x \times r_y \times r_t$ sous-régions
3. Pour chaque sous-région \mathcal{R}_i :
 - (a) Quantifier les orientations des vecteurs de gradient $d = [d_x, d_y]^T$ en n_q bins :

$$q(d) = \left\lfloor \frac{\alpha}{n_q} \right\rfloor, \quad \alpha = \arctan \left(\frac{d_y}{d_x} \right)$$

- (b) Calculer l'histogramme pondéré :

$$h^{(i)}(b) = \sum_{d|q(d)=b} ||d||_2, \quad \forall b \in [0, n_q[$$

4. Concaténer les histogrammes des régions \mathcal{R}_i :
$$h = [h^{(1)}, h^{(2)}, \dots, h^{(r_x \times r_y \times r_t)}]$$
5. Normaliser l'histogramme (L1, L2, etc.)

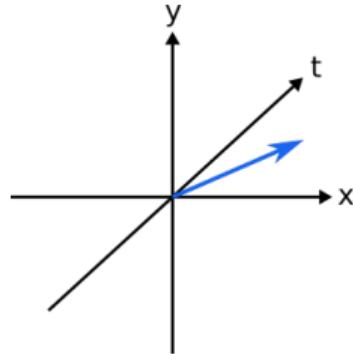
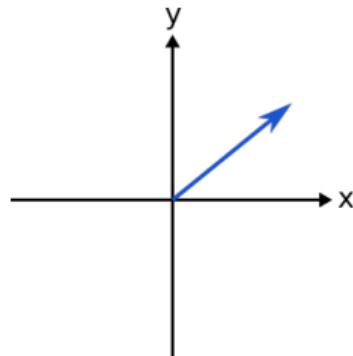


Propriété : Information d'apparence essentiellement (pas de gradient temporel)

Histogrammes de gradients spatio-temporels

Option 2 : Histogramme de gradient spatio-temporel (HOG3D) [Kläser et al., 2008]

1. Calcul du gradient spatio-temporel $\nabla I = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}]^T$
2. Découpage de la région \mathcal{R} en $r_x \times r_y \times r_t$ sous-régions
3. Pour chaque sous-région \mathcal{R}_i :
 - (a) Quantifier les orientations des vecteurs de gradient en n_q bins selon un solide régulier convexe (tétraèdre, cube, octaèdre, dodecaèdre, icosaèdre)
 - (b) Calculer l'histogramme pondéré
4. Concaténer les histogrammes des régions \mathcal{R}_i :
$$h = [h^{(1)}, h^{(2)}, \dots, h^{(r_x \times r_y \times r_t)}]$$
5. Normaliser l'histogramme (L1, L2, etc.)



Propriété : Description jointe de l'apparence et du mouvement

Histogrammes d'orientation de flux optique (HOF)

Principe : Identique au HOG cumulé

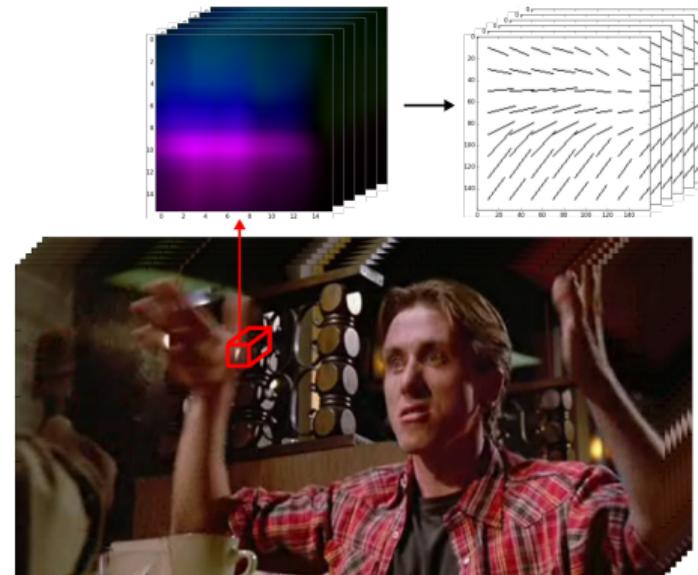
1. Calcul du flux optique $\omega = [\omega_x, \omega_y]^T$ en chaque pixel (x, y, t) de la région \mathcal{R}
2. Découpage de la région \mathcal{R} en $r_x \times r_y \times r_t$ sous-régions
3. Pour chaque sous-région \mathcal{R}_i :
 - (a) Quantifier les orientations des vecteurs de flux $\omega = [\omega_x, \omega_y]^T$ en n_q bins :

$$q(\omega) = \left\lfloor \frac{\alpha}{n_q} \right\rfloor, \quad \alpha = \arctan \left(\frac{\omega_y}{\omega_x} \right)$$

- (b) Calculer l'histogramme pondéré :

$$h^{(i)}(b) = \sum_{\omega | q(\omega)=b} ||\omega||_2, \quad \forall b \in [0, n_q[$$

4. Concaténer les histogrammes des régions \mathcal{R}_i :
$$h = [h^{(1)}, h^{(2)}, \dots, h^{(r_x \times r_y \times r_t)}]$$
5. Normaliser l'histogramme (L1, L2, etc.)



Propriété : Description du mouvement uniquement

Descripteurs spatio-temporels : Conclusion

Descripteurs spatio-temporels :

- ▶ Mouvement uniquement : HOF
- ▶ Mouvement et apparence : LBP-TOP, HOG3D
- ▶ En général, extension de méthodes spatiales à la dimension temporelle

Autres descripteurs :

- ▶ Cuboïdes : Gradients locaux + ACP [Dollár *et al.*, 2005]
- ▶ MBH : Motion boundary histogram [Wang *et al.*, 2012]
- ▶ ...

Utilisation des descripteurs : Similaire aux descripteurs d'apparence

- ▶ Matching
- ▶ Classification
- ▶ Aggrégation de descripteurs locaux
- ▶ ...

3. TP

Objectif : Reconnaissance d'actions

Contenu :

- ▶ Visualisation de points d'intérêt spatio-temporels
- ▶ Sacs de mots visuels basés sur des descripteurs HOG et HOF
- ▶ Classification
- ▶ Comparaison des descripteurs d'apparence et des descripteurs de mouvement