



# POINTS D'INTERET SPATIO-TEMPORELS POUR LA DETECTION D'OBJETS EN MOUVEMENT ET POUR LA DETECTION DE CHANGEMENTS

Alain SIMAC-LEJEUNE<sup>1,2</sup>

Michèle ROMBAUT<sup>1</sup>

Patrick LAMBERT<sup>2</sup>

<sup>1</sup>GIPSA-Lab Grenoble  
Département Images et Signal - Equipe AGPiG

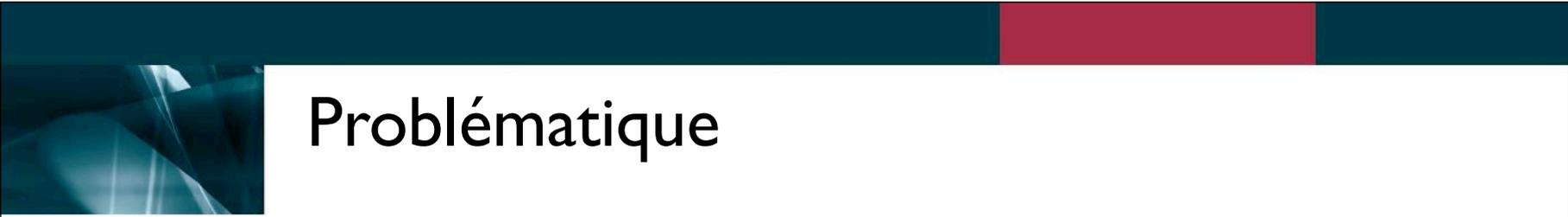
<sup>2</sup>LISTIC Annecy  
Equipe TI

MajecSTIC - 13 / 14 / 15 octobre 2010



# Contexte

Traitement de documents vidéo  
- analyse  
- indexation  
-...

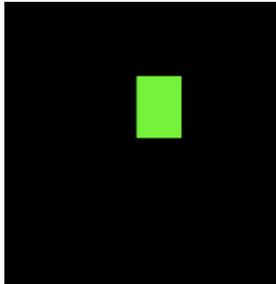


# Problématique

Extraire rapidement de l'information exploitable  
pour détecter des éléments spatio-temporels intéressants

# Base de données

Les vidéos utilisées



- 1 - Séquences de synthèse : 60 vidéos de synthèse (188x188, 150 images, 25im/s) - caméra fixe
- 2 - Séquences de sport : 40 vidéos de sauts d'athlétisme (300x300, 100 à 160 images, 25im/s) - caméra mobile
- 3 - **Film d'animation** : «Le moine et le poisson» (320x280, 8245 images, 25im/s) - caméra fixe essentiellement
- 4 - **Séquences diverses** : 60 vidéos issues de la télévision (720x576, 120 images, 25im/s) - caméra variable

**Toutes les séquences utilisées sont des plans où il n'y a pas de transitions**

**Reference 1** : Festival International du Film d'Animation - FIFA - CITIA -<http://www.annecy.org/>

**Reference 2** : Modelling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos (2009) Sophie Marat

# Sommaire

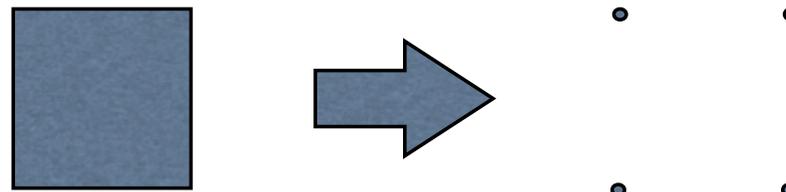
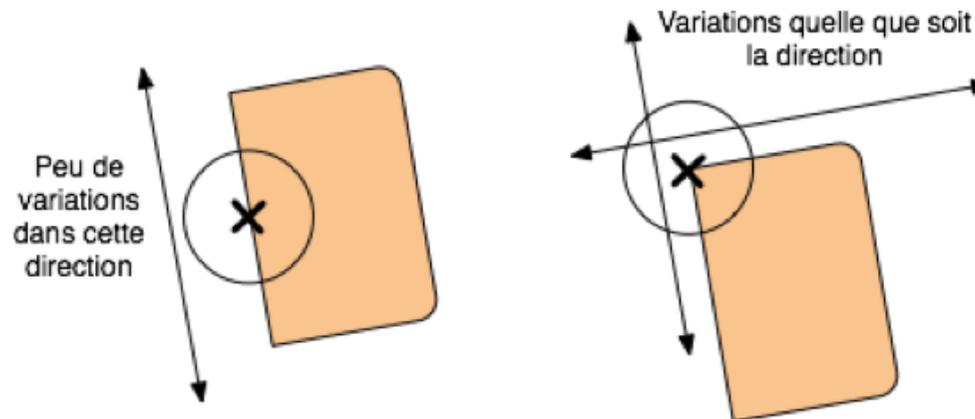
- 1- Présentation des points d'intérêt
- 2- Qualification des points d'intérêt
- 3- Détection des objets en mouvement
- 4- Détection des changements dans les mouvements
- 5- Détection des transitions
- 6- Conclusions et perspectives

# Sommaire

- 1- Présentation des points d'intérêt
- 2- Qualification des points d'intérêt
- 3- Détection des objets en mouvement
- 4- Détection des changements dans les mouvements
- 5- Détection des transitions
- 6- Conclusions et perspectives

# Point d'intérêt spatial

Résume l'information en quelques points



# Définition des Points d'Intérêt Spatiaux (SIP)

coins dans l'image

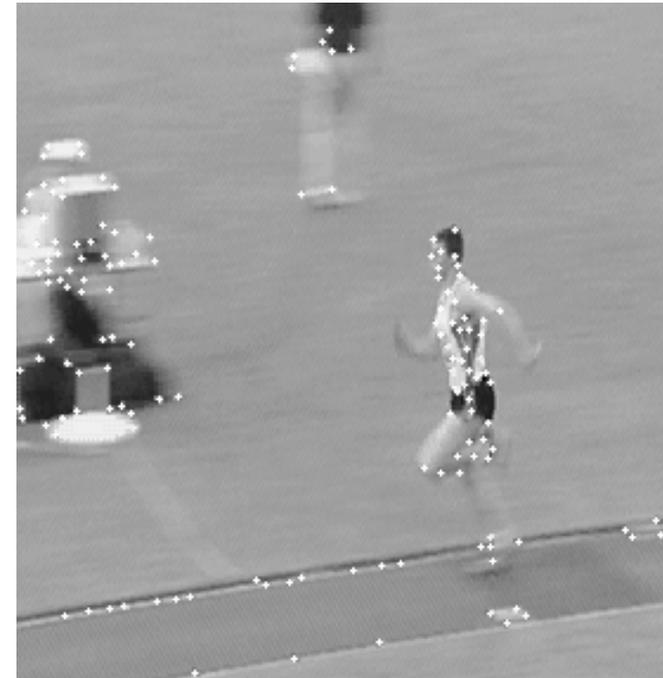
Méthode d'extraction

Matrice du filtre Hessian-Laplace

$$M(I(x, y)) = G(x, y) \otimes \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{pmatrix}$$

Extrema locaux positifs

$$R(x, y) = \det(M) - k \times \text{tr}(M)^2$$



On considère les pixels  $I(x,y)$  /  $G(x,y)$  gaussienne de variance  $\sigma_s$

Reference : A combined corner and edge detector (1988) Chris Harris and Mike Stephens

# Définition des Points d'Intérêt Spatio-Temporels (STIP)

coins dans l'image qui ont un mouvement désordonné dans le temps

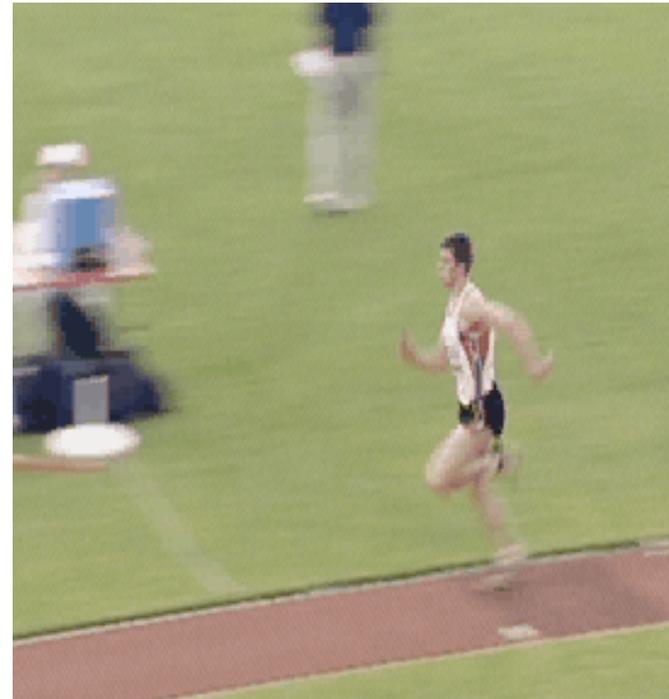
## Méthode d'extraction

Matrice du filtre Hessian-Laplace

$$M(x, y, t) = G \otimes \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial x \partial t} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} & \frac{\partial^2 I}{\partial y \partial t} \\ \frac{\partial^2 I}{\partial x \partial t} & \frac{\partial^2 I}{\partial y \partial t} & \frac{\partial^2 I}{\partial t^2} \end{pmatrix}$$

Extremums locaux positifs

$$R(x, y) = \det(M) - k \times \text{tr}(M)^3$$



On considère les pixels  $I(x,y,t) / G(x,y,t)$  gaussiennes de variance  $\sigma_s \sigma_t$

Reference : On space-time interest points (2005) Ivan Laptev

# Sommaire

- 1- Présentation des points d'intérêt
- 2- Qualification des points d'intérêt
- 3- Détection des objets en mouvement
- 4- Détection des changements dans les mouvements
- 5- Détection des transitions
- 6- Conclusions et perspectives

# Influence des variances

Nombre moyen de STIP généré par image

Précision correspondant au pourcentage de points sur un objet en mouvement

$\sigma_s$	$\sigma_t$	STIP/image	Précision
1.5	1.5	68	96,2%
1.2	1.5	213	84,4%
2.0	1.5	9	100%
1.5	1.2	191	85,7%
1.5	2.0	19	99,7%
1.2	1.2	521	71,9%
2.0	2.0	1	100%

Test - 20 séquences (8 sauts, 2 animations, 10 divers)

$k = 0,04$  et  $S = 150$

$\sigma_s$  : influence sur la gaussienne spatiale = nombre de pixels pris en compte

$\sigma_t$  : influence sur la gaussienne temporelle = nombre d'images prises en compte

$k$  : coefficient fixé empiriquement à 0,04 (littérature) et  $S$  le seuil fixé à 150

# Influence de l'orientation de la caméra

Nombre moyen de STIP généré par image

Précision correspondant au pourcentage de points sur un objet en mouvement

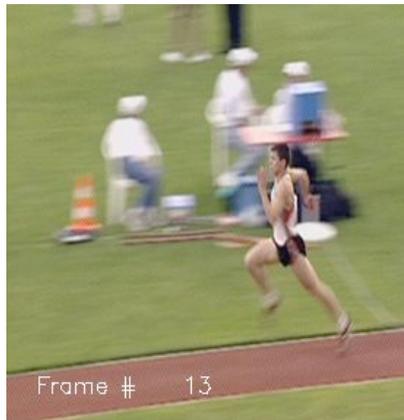
	STIP/frame	Bonne détection
Caméra de face	76	91%
Caméra de profil	91	99%
Caméra statique	82	93%
Caméra mobile	83	98%

Test - séquences de sauts d'athlétisme  
 $k = 0,04$ ,  $\sigma_s = \sigma_t = 1.5$  et  $S = 150$

Globalement, les points sont davantage situées sur le personnage en mouvement dans le cas de caméra de profil que dans le cas de caméra de face

Les caméras mobiles permettent d'avoir une meilleur détection des mouvements car le mouvement de la caméra permet de «filtrer» les points générés hors d'objet en mouvement (bruit)

# Influence du contraste et du bruit



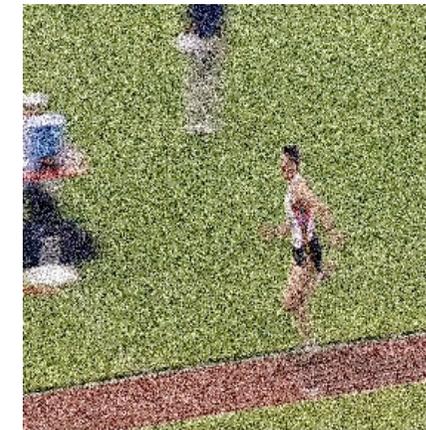
Référentiel



Augmentation du contraste



Diminution du contraste



Ajout de bruit

Contraste	STIP/image
25%	0
50%	< 1
75%	11
100% reference	68
125%	155
150%	203
175%	231

a) Influence du contraste

Bruit		STIP/image
Puissance	Intensité	
0	0	68
20	20	69
20	50	70
20	70	72
20	100	74
50	20	83
50	50	119
50	70	153
50	100	224

b) Influence du bruit

Test - séquences d'athlétisme  
 $k = 0,04$ ,  $\sigma_s = \sigma_t = 1.5$  et  $S = 150$

Problème :

les points supplémentaires  
 ne sont pas forcément  
 pertinents

# Sommaire

- 1- Présentation des points d'intérêt
- 2- Qualification des points d'intérêt
- 3- **Détection des objets en mouvement**
- 4- Détection des changements dans les mouvements
- 5- Détection des transitions
- 6- Conclusions et perspectives

# Détection des objets en mouvement

Objet d'intérêt en mouvement : fort contenu sémantique



Caméra fixe



Caméra mobile

# Détection des objets en mouvement

Séquences de synthèse

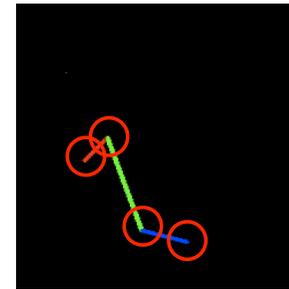
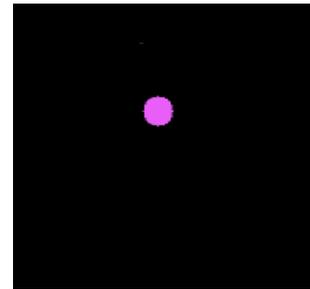
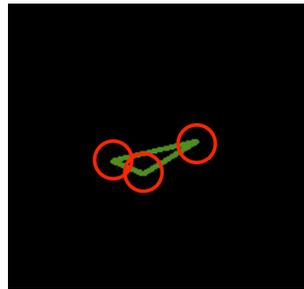
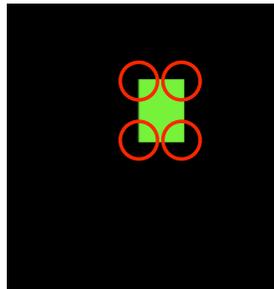
## REGLES DE DETECTION

- vrai positif : au moins un point sur un objet en mouvement
- faux positif : au moins un point sur un objet qui n'est pas en mouvement
- faux négatif : aucun point sur un objet en mouvement

	cercles	carrés	triangles	lignes brises
Précision	N/A	1.00	1.00	0.99
Rappel	N/A	1.00	1.00	0.99

séquences de synthèse

$k = 0,04, \sigma_s, \sigma_t = 1.5$  et  $S = 120$



# Séquences réelles

## REGLES DE DETECTION

- vrai positif : au moins un point sur un objet en mouvement
- faux positif : au moins un point sur un objet qui n'est pas en mouvement
- faux négatif : aucun point sur un objet en mouvement

Un point est considéré comme appartenant à un objet en mouvement si le point est dans la 'box englobante' de l'objet en mouvement

	Précision	Rappel
<b>Films d'animation</b>	0.99	0.91
<b>Vidéos d'athlétisme</b>	0.93	0.95

séquences d'athlétisme et du film "Le Moine et le Poisson"

$k = 0,04, \sigma_s, \sigma_t = 1.5$  et  $S = 120$

# Sommaire

- 1- Présentation des points d'intérêt
- 2- Qualification des points d'intérêt
- 3- Détection des objets en mouvement
- 4- Détection des changements dans les mouvements
- 5- Détection des transitions
- 6- Conclusions et perspectives

# Détection de changements dans les mouvements

Changement à 90° et à 180° sur les vidéos de synthèse

METHODE DE DETECTION : suivi de l'activité

Définition de l'activité :  $0.8 \times \text{nombreDeSTIP}(t - \alpha) \leq \text{nombreDeSTIP}(t) \leq 0.8 \times \text{nombreDeSTIP}(t + \alpha)$

alpha : nombre d'images - correspond à  $\sigma_t$

Qualification de la méthode sur les données de synthèse

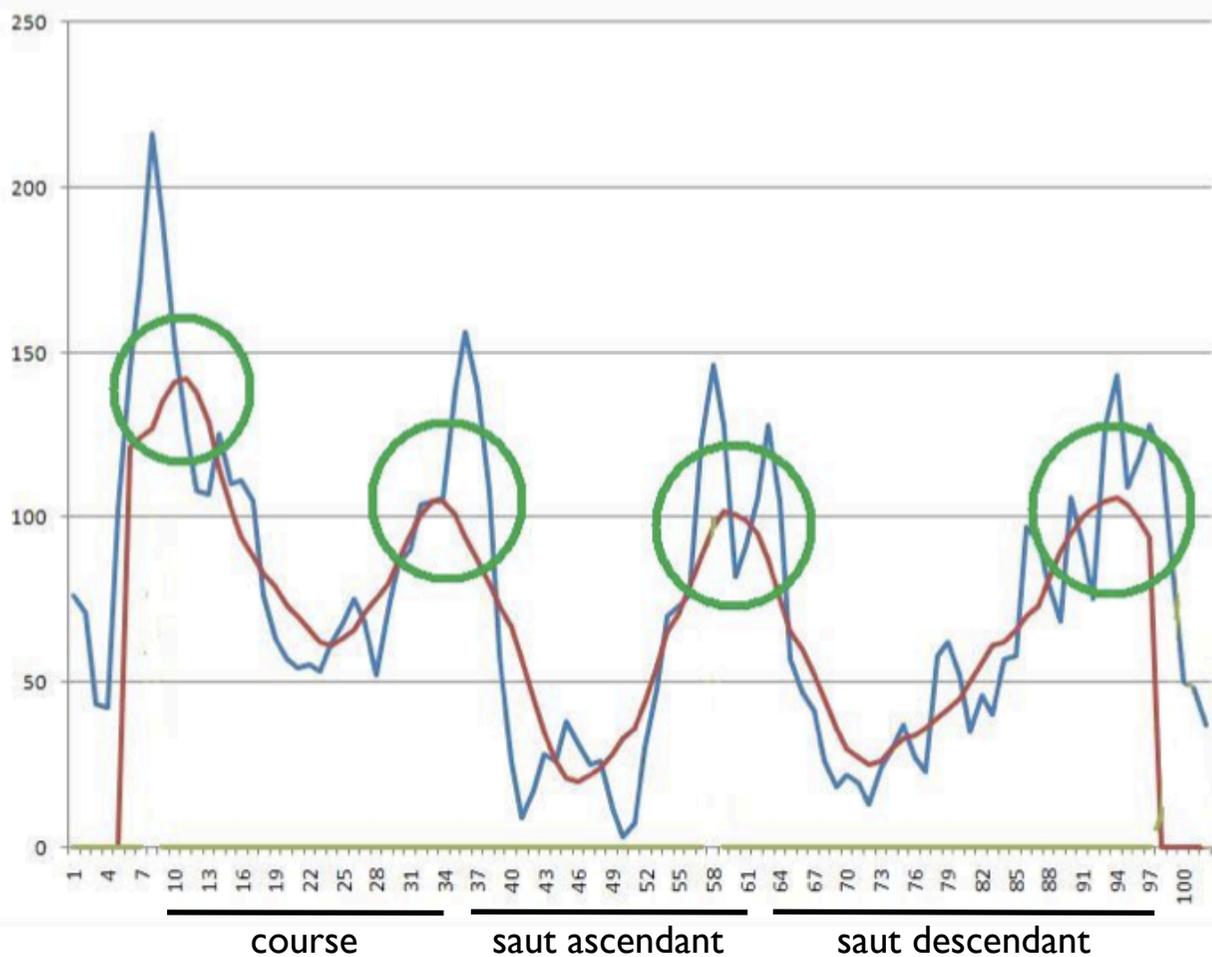
	Précision	Rappel	Tolérance
90°	1.00	1.00	±3frames
Inversion	1.00	1.00	±3frames
Tous	1.00	0.98	±3frames

séquences de synthèse

$k = 0,04, \sigma_s = \sigma_t = 1.5$  et  $S = 120$

# Détection de changements dans les mouvements

Utilisation sur un saut d'athlétisme



- 2 courbes  
nombre de STIP
- bleue : instantanée
  - rouge : lissée

Exemple

Saut en longueur  
4 points détectés :

- début course
- appel
- phase ASC/DES
- athlète qui se relève

# Détection de changements dans les mouvements

Utilisation sur différentes séquences

	Précision	Rappel	Tolrance
Saut en longueur	0.93	0.92	±3frames
Saut en hauteur	0.92	0.88	±3frames
Triple saut	0.81	0.71	±5frames
Saut à la perche	0.84	0.85	±10frames
Hand-ball	0.76	0.86	±3frames
Circulation automobile	0.97	0.93	±3frames

20 séquences de saut, 8 séquences de hand-ball et 8 de surveillance routière

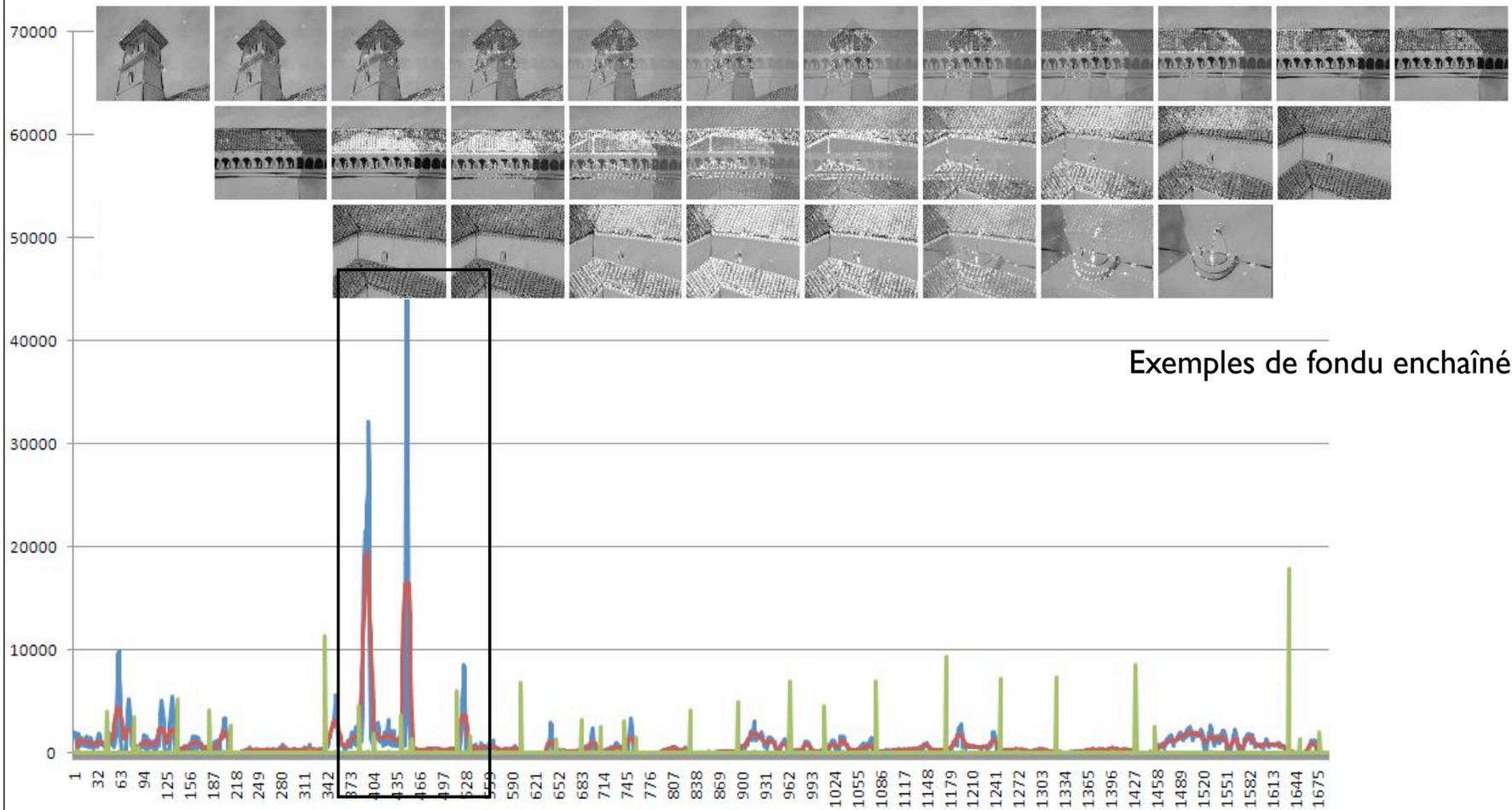
$k = 0,04$ ,  $\sigma_s = \sigma_t = 1.5$  et  $S = 120$

Pour chaque changement  $G$  d'une séquence  $S$ , on disposait d'une **vérité terrain annotée manuellement** telle que :  $G\_S(1) = 15$  /  $G\_S(2) = 37...$

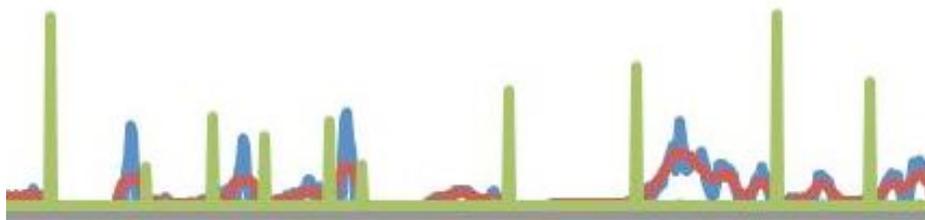
# Sommaire

- 1- Présentation des points d'intérêt
- 2- Qualification des points d'intérêt
- 3- Détection des objets en mouvement
- 4- Détection des changements dans les mouvements
- 5- Détection des transitions
- 6- Conclusions et perspectives

# Détection des transitions entre plans



# Détection de transitions entre plans



Transitions détectées (courbe verte)



Vérité terrain

En fonction du type de transitions : précision allant de 80 à 100% (0.8-1.0), rappel moyen de 90% (0.9)

2 types de transitions : coupe / fondu enchaîné

=> coupe très bien détectée / plus les fondus sont courts mieux ils sont détectés

# Conclusions et perspectives

## 1. La détection d'objets en mouvement

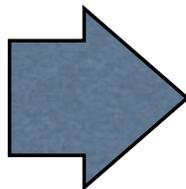
ciblage d'objets d'intérêt potentiels => patch/segmentation

## 2. La détection de changements

segmentation en phases cohérentes en terme de mouvement

## 3 La détection de transitions

segmentation en plan, même pour des transitions complexes (fondus enchaînés)



### POINTS D'INTERET SPATIO-TEMPORELS

structuration et focalisation pour des documents vidéos  
rapidité de calcul (24im/sec sur C2Duo 2,6Ghz / 4GoRam)

# Remerciements

Rhône-Alpes <sup>Région</sup>

CLUSTERS  
DE RECHERCHE  
RHÔNE-ALPES

ISLE Informatique, Signal,  
Logiciel Embarqué

## Merci pour votre attention

