

# Vision par ordinateur du mouvement

## Projet Vidéo

Dr. Jean-Christophe CEXUS  
-  
ENSTA-Bretagne

Laboratoire STIC/REMS - LabSTICC UMR CNRS 6286

[Jean-christophe.cexus@ensta-bretagne.fr](mailto:Jean-christophe.cexus@ensta-bretagne.fr)  
Bureau : E103

Janvier 2020

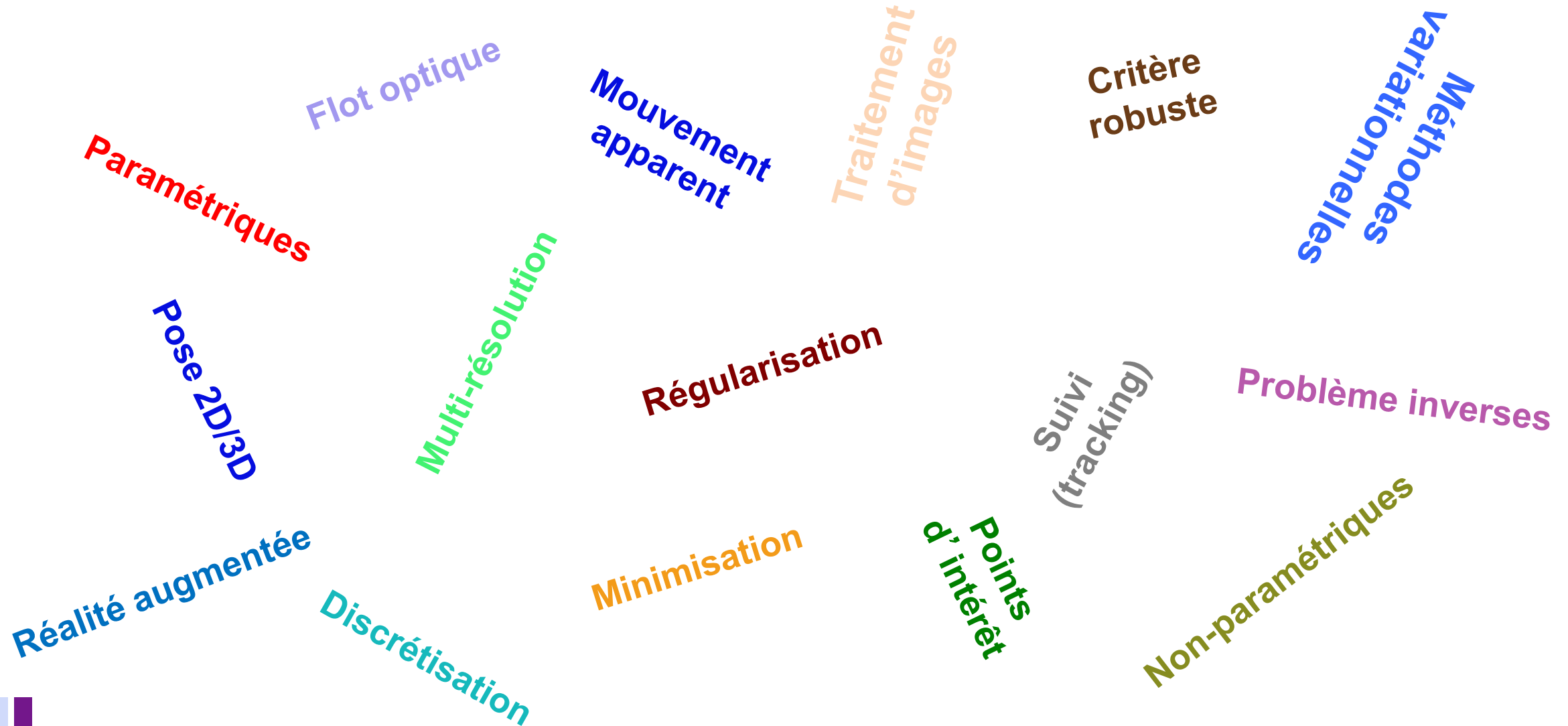
## Planning du cours

### Projet Vidéo :

#### Autour de la vidéo : vision par ordinateur du mouvement

- Introduction
- Projet vidéo
  - Etude d'un thème précis en groupe
  - Restitution :
    - Présentation type power-point
    - Présentation d'un démonstrateur (*preuve de concept*)

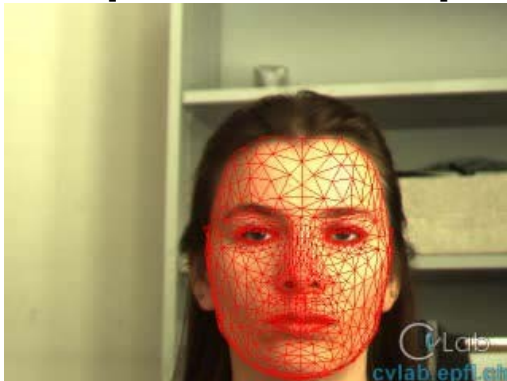
## > Introduction : applications et enjeux



## > Introduction : applications et enjeux

### Estimation/suivi de pose – Modélisation 3D (Réalité virtuelle – Réalité augmentée)

[Vacchetti *et al.*, 2003]



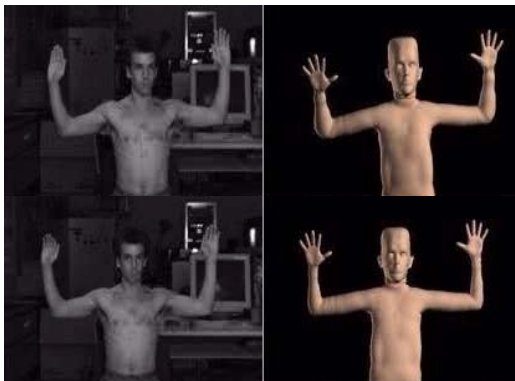
[Comport *et al.*, 2008]



[Hugues *et al.*, 2011]



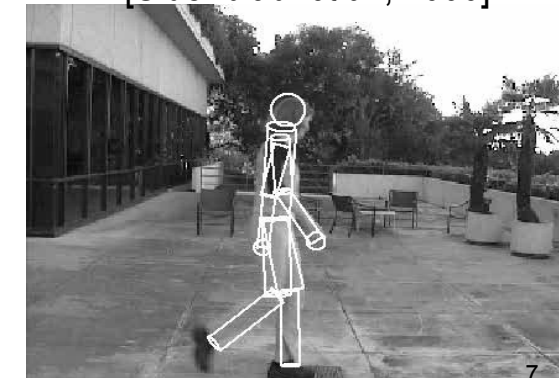
[Plaenkers & Fua, 2003]



[Sudderth *et al.*, 2004]



[Sidenblad *et al.*, 2000]



## > Introduction : applications et enjeux

### Détection / suivi d'objets

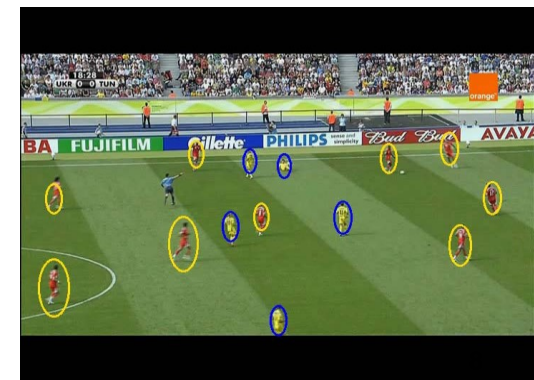
[Viola and Jones, 2001]



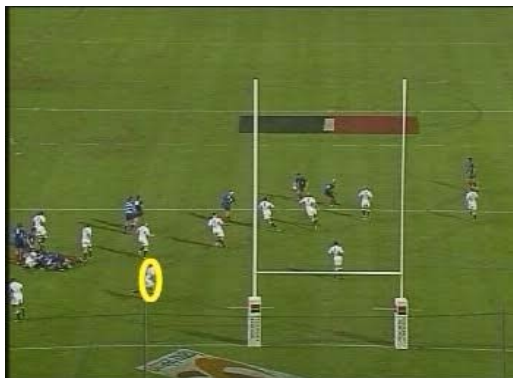
[Okuma *et al.*, 2004]



[Lehuger *et al.*, 2007]



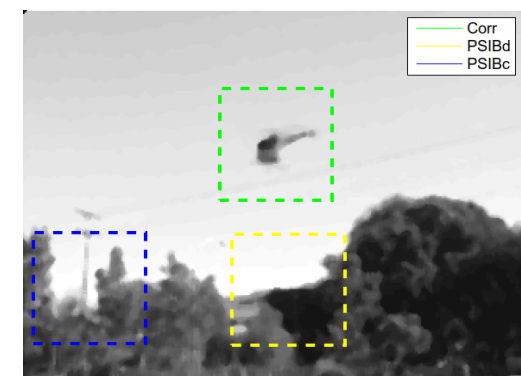
[Gengembre and Pérez, 2006]



[Pérez *et al.*, 2002]



[Farah *et al.*, 2009 (master II)]





## > Introduction : applications et enjeux

### Recherche / résumé dans une vidéo

**Film prétraité** : extraction des points d'intérêts, caractérisation,

**Requête** : recherche d'un '*objet*' (puis extension par suivi)

**Résultat** : recherche d'autres instances de l'objet dans le film (ou d'autres vidéos)

**Résumés** : extraction d'une ou plusieurs images clés par séquence

[Vermaak et al., 2002]



[Sivic et al., 2004]

"VideoGoogle" at Oxford Visual Geometry Group



[N. Peyrard et al., 2005]

## > Introduction : applications et enjeux

Post-traitements : élimination 'objets, stabilisation, ...

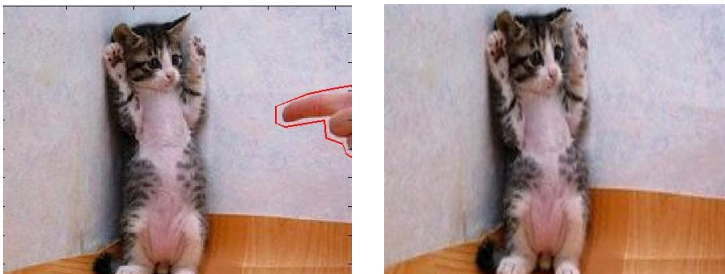
**Outils:** estimation du mouvement, suivi, synthèse de texture par l'exemple

**Applications :** restauration, corrections, effets spéciaux (*resize, extraction, compresion, ...*)

[Grossauer et al., 2006]



[Wexler et al., 2007]



[Berthomier et al., 2014 (projet indus)]



[Rubinstein, Shamir et al., 2008]



## > Introduction : applications et enjeux

**Post-traitements : analyse de gestes sportifs, analyse de la marche**

**Outils:** estimation du mouvement, suivi, synthèse de texture par l'exemple

**Applications :** vitesse, trajectoire par suivi, alignement de séquences,



CV Lab at EPFL with DartFish  
[Fua *et al.*, 2003]



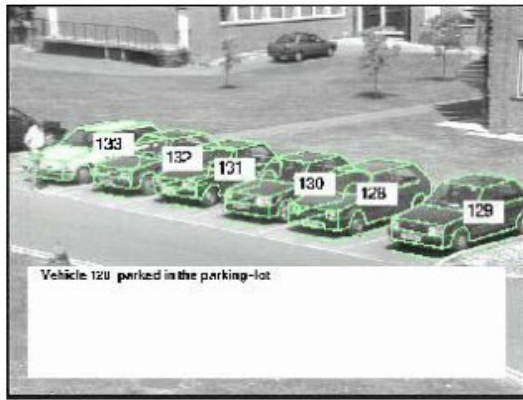
DartFish's SimulCam  
[Dexter *et al.*, 2009]



## > Introduction : applications et enjeux

**Autres applications : larges spectres**  
**Autres capteurs (optronique, médicale...)**

Surveillance de zone  
 (vidéo surveillance)



[Kilger *et al.*, 1992]



Domaine militaire  
 (infra-rouge, radar)



Navigation robotique (AUV)



[Biswas *et al.*, 2019]

Poursuite automatique  
 (pistage)



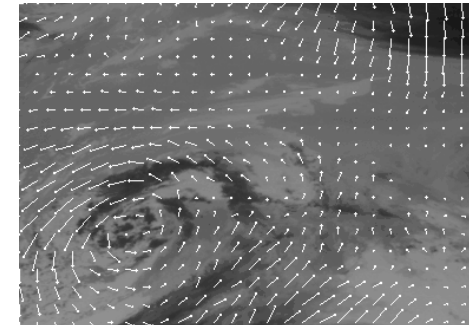
[Chen Z. *et al.*, 2019]

Surveillance de feu



[Antoun *et al.*, 2019]

Domaine météorologique



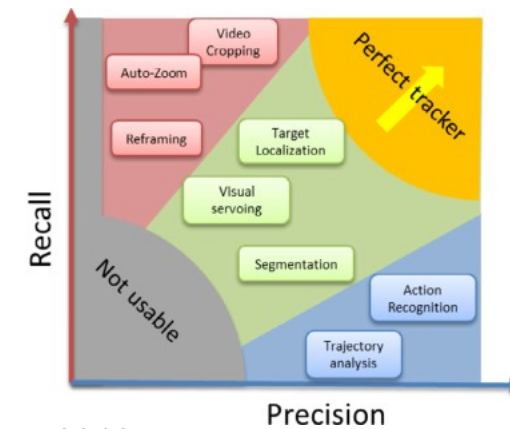
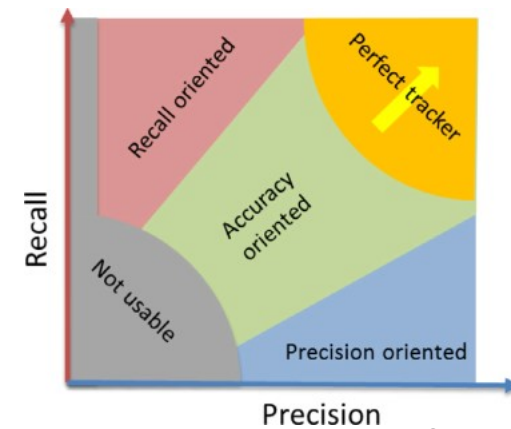
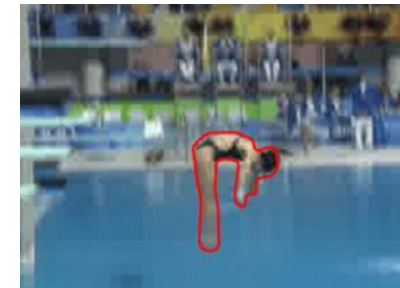
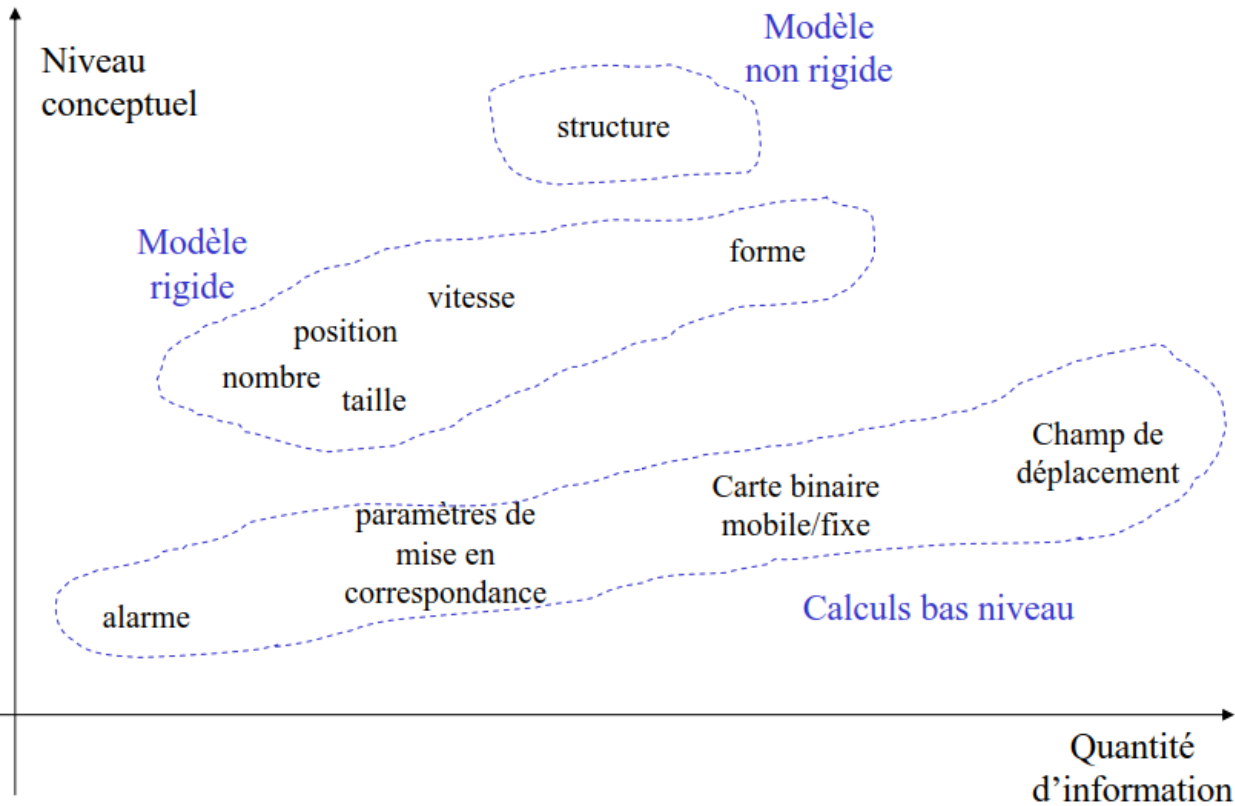
[Corpetti *et al.*, 2001]

Domaine médicale  
 (scanner)

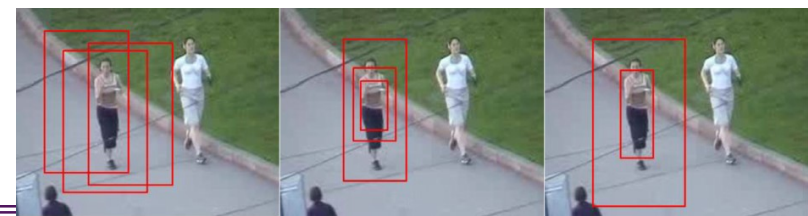


## > Introduction : applications et enjeux

Quelle information extraire ?

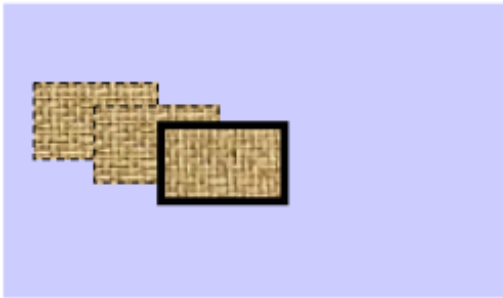


[Crivelli et al., 2016]



## > Introduction : applications et enjeux

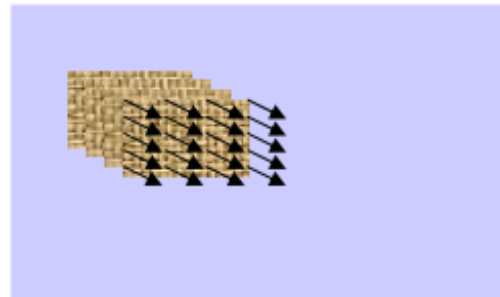
### Les grandes problématiques



#### DETECTION

Objectif : identifier dans chaque image les pixels appartenant à des objets mobiles (identiques).

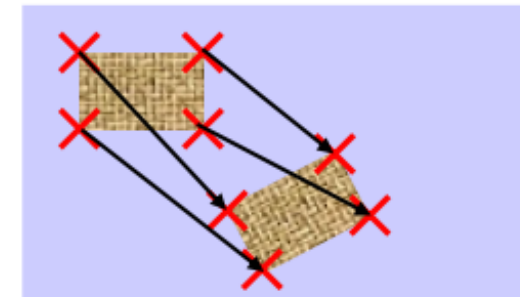
- Continuité temporelle du mvt
- Mouvement de la caméra (ou pas)



#### ESTIMATION

Objectif : calculer le mouvement apparent (vitesse instantanée) de chaque pixel.

- Continuité temporelle
- Problématique des petits déplacement vs grands déplacements



#### POURSUITE (suivi)

Objectif : apparier certaines structures spatiales pour chaque couple d'images.

- Discontinuité temporelle
- Problématique des objets rigides vs objets non-rigides

## > Focus sur la détection

### Détection du mouvement

- Détection de changement (recherche des singularités dans les séries temporelles)
- Problématique du fond (représentation, modélisation, histogramme)
- Des cas plus ou moins compliqué, complexité des algorithmes



### Contexte

- Camera mobile (modélisation du mvt de la caméra)
- Objets mobiles (son mvt) et déformables (ou pas)

### Notions de Fond et d'objets

- Fond (Background) : Scène fixe (mais pas forcément statique !)
- Objets (Foreground) : Objets mobiles (mais pas en permanence !)

### Notion & Concepts

- Traitement de séries temporelles (gradient temporelle, statistiques temporelles),
- Classification Foreground / Background (binarisation dans des espaces de représentation type HSV, YUV ...)
- Estimation robuste, non-stationnaire, modélisation du fond , problématique de régularisation (optimisation / bruit)
- Complexité : compromis coût de calcul (temps, mémoire), algorithmique,



Vidéo



Fond



Objets



## > Focus sur l'estimation

### Estimation du mouvement

- Mise en correspondance des pixels entre deux images (recalage)
- Quels éléments d'une image correspondent aux éléments de l'image suivante ?



[Odobez et al,1997]

### Deux approches possibles

- **Petits déplacements** : méthodes différentielles basées sur l'analyse du mouvement apparent
- **Grands déplacements** : méthodes (très nombreuses) basées sur l'extraction de primitives (point d'intérêt, structures des objets connus ...) → Problématique avec la poursuite

### Notions de mouvement apparent

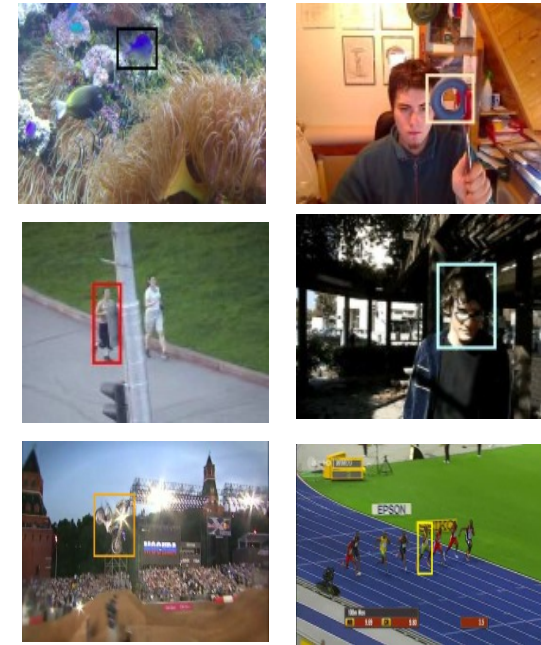
- **Hypothèse** : malgré le mouvement, l'intensité lumineuse de l'objet reste constante

### Notion & Concepts

- Équation de Contrainte du Mouvement Apparent (ECMA)
- Notion de base du flux optique (estimation dense)
- Estimation du mouvement dominant de la caméra si elle est en mouvement (*pan, tilt, roll, zoom, ...*)



$$I(x, y, t) = I(x + dx, y + dy, t + 1)$$



## > Focus sur la poursuite

### Poursuite – Suivi - Tracking

- Algorithme de suivi est généralement composé de deux étapes :
  - **Prédiction** : hypothèse initiale de localisation de l'objet basé sur un *modèle dynamique* (cinématique de l'objet, de la caméra)
  - **Détection** : localiser l'objet à partir d'*observations* extraites de l'images (*modèle d'apparence*) → mesure de (di)similarité
- Complexité algorithmiques, coût de calcul

### Contexte

- Camera fixe ou mobile
- Mouvements des objets / Déformation des objets

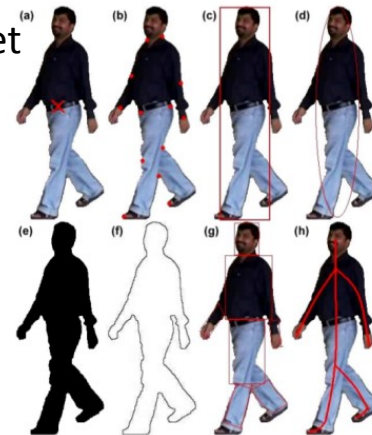
### Objectifs

- Suivre et localiser un objet défini initialement (*comment* ? manuellement ou non) en estimant dans chaque image le support spatial (*lequel* ?) englobant l'objet

### Notion & Concepts

- Support spatial de l'objet : flexible, fidélité, complexité, **invariances**, ...
- Challenges du suivi vis-à-vis de la déformation, complexité du fond (variabilité, illumination, faible contraste), occultations, mouvement de la caméra (flou de bougé – restauration)
- Notion de filtrage prédictif (filtre de Kalman ou filtre particulaire)

Support spatial de l'objet



[Jalal et al., 2012]

## > Focus sur la poursuite

### Mesure de similarité

- Les mesures de (di)similarité globale s'applique entre deux vecteurs  $T$  et  $X$  de même dimension  $n$ , dont l'un représente le modèle (**template**) de l'objet sous la forme (en général) d'une image rectangulaire, et l'autre une image extraite de l'image courante correspondant à une hypothèse de localisation.
- L'algorithme de détection consiste à parcourir un espace d'hypothèses de localisation  $\{X^h\}_{h \in H}$  pour trouver celle qui minimise (ou maximise) la (di)similarité :

$$I = \arg \min_{h \in H} D_k(T, X^h)$$

### Exemple de la distance de Minkowski (min)

$$D_k(T, X) = \sum_{i \leq n} |T_i - X_i|^k$$

- Sensible aux déformations, aux variations d'illuminations
- Facile à mettre en œuvre



### Exemple du coefficient de corrélation normalisé (max)

$$D_k(T, X) = \frac{\sum_{i \leq n} (T_i - \tilde{T}_i) ((X_i - \tilde{X}_i))}{\sqrt{\sum_{i \leq n} (T_i - \tilde{T}_i)^2} \sqrt{\sum_{i \leq n} (X_i - \tilde{X}_i)^2}}$$

- Sensible aux déformations
- robuste aux variations d'illuminations

## > Focus sur la poursuite

### Poursuite et recalage dans le cas de grands déplacements

Lorsque l'amplitude du déplacement est grande, l'estimation directe par les méthodes d'appariement ou les méthodes différentielles sont souvent vouées à l'échec principalement pour deux raisons :

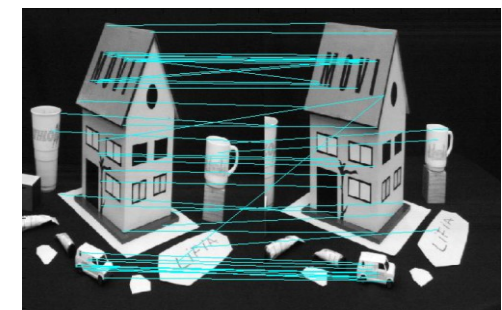
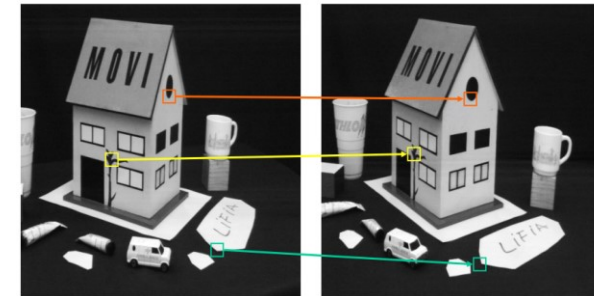
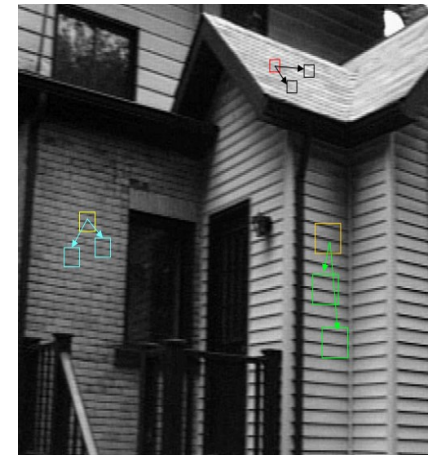
- Le déplacement apparent peut être supérieur aux périodes spatiales présentes dans les images (aliasing temporel).
- La scène peut-être soumise à des déformations géométriques complexes qui invalident l'hypothèse de translation locale.

### Conséquences :

- Il est important de disposer **de descripteurs qui soit le plus invariants** possible aux transformations géométriques : rotation, homothétie, transformation affine.  
→ Cela passe très souvent par un calcul *multi-échelle*.
- Les descripteurs fournis ont un caractère beaucoup moins local, voire global, et on ne dispose plus d'un champ dense (flot optique), mais au mieux d'un champ épars.

### Méthodes :

- Transformation globale : méthodes fréquentielles (Transformée de Fourier)
- Appariement par descripteurs : points anguleux (intérêt, saillant...) Harris, SIFT ...



Descripteur SIFT :  
histogramme d'orientation



## > Le projet vidéo : Déroulement

- **Etude d'un thème précis avec une réalisation (démonstrateur)**
  - Tracking par Flot Optique (FO)
  - Tracking par Mean-Shift (MS)
- **Deux salles informatiques disponibles durant 3 créneaux de 4 heures**
  - F016 : 16 étudiants → 4 groupes de 4 étudiants
  - F015 : 17 étudiants → 3 groupes de 4 étudiants et 1 groupes de 5

Dates	Horaire	Séquencement des 3 séances
Vendredi 24 janvier	08h10 – 12h15	Introduction – Constitution du groupe - Découverte du thème
Lundi 17 février	13h30 – 17h30	Poursuite de l'étude thématique
Jeudi 20 février	08h10 – 12h15	Finalisation – Restitution des groupes (ppt & démonstrateur)

- **Constitution des groupes de 4 étudiants :** (mélange des différents cursus par groupe)  
(constitution des groupes avant la fin de la première séance avec le professeur)

Dans chaque salle

Groupe : thème	Etudiant 1	Etudiant 2	Etudiant 3	Etudiant 4	Etudiant 5
Groupe 1 : <b>FO</b> (1)	<i>Joe (FIPA)</i>	<i>William (PSO)</i>	<i>Jack (ROB)</i>	<i>Averell (FIPA)</i>	<i>Luke (S-W)</i>
Groupe 2 : <b>MS</b> (2)					-
Groupe 3 : <b>FO</b> (3)					-
Groupe 4 : <b>MS</b> (4)					-

## > Le projet vidéo : Objectifs pour chaque groupe

### ➤ Réalisation d'un support de type PowerPoint contenant les points suivants : (~10 à 15 planches)

- Contexte et objectifs
- Hypothèse(S) de travail (par exemple webcam fixe, ...)
- Explications théoriques (un peu de mathématiques)
- Présentation du démonstrateur : structures du programmes via un diagramme
- Simulations et résultats : avantages et inconvénients
- Conclusions vers quelques extensions ...
- Références – liens internet

Dépôt sous Moodle  
20 février  
dans la journée

### ➤ Démonstrateur (petit – preuve de concept)

- Implémentation en Python ou Matlab
- Obligation de mise en œuvre d'une **webcam** (ou camera sur ordinateur)
- Possible d'utiliser des bibliothèques se trouvant sur internet : **citer vos sources**

### ➤ Restitution (~20 min) pour chaque groupe : jeudi 20 février

- Présentation d'un groupe à un autre groupe : **FO(i) → MS(i+1)** puis **MS(i+1) → FO(i)** avec *i=1 ou 3*
- Présentation : 10 min, Démonstrateur : 5 min, Questions/Réponses : 5 min

Gr. présente à Gr.		Début	Fin
FO(1)	MS(2)	11h10	11h30
FO(3)	MS(4)	11h10	11h30
MS(2)	FO(1)	11h40	12h00
MS(4)	FO(3)	11h40	12h00

Dans les deux salles  
J.-C.Cexus & A. Toumi

## > Le suivi d'objet dans une vidéo : flot optique

Le calcul d'un mouvement apparent **global** (mise en correspondance) entre deux images correspond à l'estimation des paramètres d'une transformation affectant tous les points de l'image : translation, rotation, homothétie, affinité ...

Le calcul du mouvement apparent **local** consiste à associer à chaque pixel  $(x, y, t)$  de  $I$  un vecteur  $(v_x(t), v_y(t))$  représentant la vitesse apparente du pixel  $(x, y)$  à l'instant  $t$ .

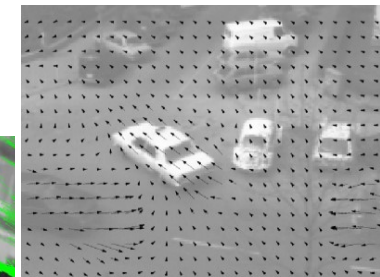
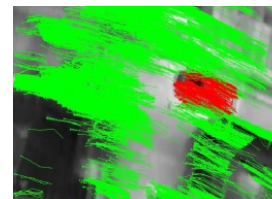
→ Calcul du flot optique (= Champ de mouvement apparent / Appariement local des points)

### Flot optique

- Champ dense de vitesse décrivant le mouvement apparent des motifs d'intensités de l'image sous l'hypothèse d'illumination constante et de petits mouvements.
- Problème inverse mal posé ? Estimation des caractéristiques physiques de l'environnement à partir d'informations ne permettant que de les retrouver partiellement
- Nécessaire d'utiliser des information complémentaire (a priori) : régularisation spatiale et temporelle

### Deux approches basées sur les variations temporelles de $I(x, y, t)$

- Flot optique approche de Lucas-Kanade (1981)
- Flot optique approche de Horn-Schunk (1981)



gradient spatial

Vitesses  $(x, y)$  inconnues

$$\nabla I \cdot v + \frac{\partial I}{\partial t} = 0$$

gradient temporel

→ Application à la poursuite

## > Le suivi d'objet dans une vidéo : Mean-Shift

**Formulation du suivi** entre deux images est basée sur la recherche des paramètres de déformation permettant de conserver les descripteurs de la région (de l'objet) d'une image à une autre : principe des mesures de (di)similarité, filtrage particulière ou encore des approches basées sur le Mean-Shift ...

### Mean-Shift

- L'objet (la région) définie par une ellipse (un noyau).
- Description de la région basée sur un histogramme (des couleurs)
- Principe de base : les paramètres de l'ellipse étant connus à l'instant  $t$ , il s'agit de trouver les paramètres à  $t+1$  de telle manière à minimiser les distances entre les histogrammes

$$I = \arg \min_{h \in H} \mathbf{B}(H(X, I_t), H(X^h, I_{t+1}))$$

- Avec  $\mathbf{B}(. , .)$  classiquement la distance (ou index) de Bhattacharyya permettant de mesurer **la similarité entre deux distributions** (ici des histogrammes normalisés).

### Avantages / inconvénients :

- Robuste à tout type de déformation
- Sensible aux variations d'illumination
- Peu discriminant sur le plan géométrique



[Comaniciu *et al.*, 2003]

→ Application à la poursuite



## Le projet vidéo : si nous avons à évaluer l'étude – Théoriquement

### ➤ **Formulation et modélisation du problème traité**

- **A.** Formule le problème à travers un modèle dont toutes les hypothèses sont clairement définies.
- **B.** Formule le problème à travers un modèle dont la plupart des hypothèses sont spécifiées.
- **C.** Introduit le modèle mathématique associé au problème traité
- **D.** Le lien entre le problème traité et le modèle considéré est confus

### ➤ **Méthodes de résolution du problème traité**

- **A.** Explique les points-clés associés à la résolution théorique du problème
- **B.** Fournit des éléments d'explication sur la résolution théorique du problème
- **C.** Décrit la solution théorique de résolution du problème
- **D.** Ne fournit pas d'éléments théoriques sur la résolution du problème

### ➤ **Un challenge entre les différents groupes travaillant sur le même thème !**

## > Le projet vidéo : si nous avons à évaluer l'étude – Expérimentalement

### ➤ Mise en œuvre algorithmique

- **A.** Etablit et explique les liens entre le modèle considéré et sa mise en œuvre algorithmique, y compris vis-à-vis des hypothèses-clés du modèle et d'autres algorithmes classiques
- **B.** Décrit correctement la mise en œuvre algorithmique et repère les liens entre le modèle considéré et les principales étapes algorithmiques
- **C.** Décrit correctement l'algorithme proposé pour résoudre le problème traité
- **D.** Ne décrit pas un algorithme permettant de résoudre le problème de traité

### ➤ Evaluation expérimentale

- **A.** Synthétise l'évaluation expérimentale réalisée et en fournit une analyse critique vis-à-vis de différents modèles et/ou solutions algorithmiques
- **B.** Synthétise l'évaluation expérimentale réalisée (plan d'expérience, critère d'évaluation)
- **C.** Fournit des éléments d'évaluation expérimentale de la méthode proposée
- **D.** Ne décrit pas un algorithme permettant de résoudre le problème de traité

## > Quelques éléments de bibliographie

### Bibliographie - Détection

A. Elgammal, *et al.*

Non-parametric Model for Background Substruction  
Proc. of ICCV '99 FRAME-RATE Workshop (1999)

C. Stauffer & C. Grimson

Learning patterns of activity using real-time tracking.  
IEEE Trans. on PAMI 22(8), 747-757. (2000)

P. Power & J. Schonees

Understanding background mixture models for foreground segmentation.  
Imaging and Vision Computing New Zealand, Auckland (2002)

B.D. Lucas & T. Kanade

An iterative image registration technique with an application to stereo vision  
International Journal of Computer Vision and Artificial Intelligence 674-679 (1981)

B.K.P Horn & B. Schunck

Determining Optical Flow »  
Artificial Intelligence 23, 185-203 (1981)

<http://petercorke.com/wordpress/books/book>

### Bibliographie - Suivi

A.S. Jalal & V. Singh

The State-of-the-Art in Visual Object Tracking  
Informatica 36 (2012) 227-248

M. Kristan *et al.*

The Visual Object Tracking VOT2014 challenge results  
Visual Object Tracking Workshop 2014, 2014

G. Welsh & G. Bishop

An Introduction to the Kalman Filter  
Tutorial of ACM SIGGRAPH (2001)

D Comaniciu, V. Ramesh & P. Meer

Kernel-based object tracking  
Pattern Analysis and Machine Intelligence,  
25(5), 564-575 (2003)

D.H. Ballard & C.M Brown Computer Vision, Prentice Hall (1982)

R. Jain, *et al.* Machine Vision, McGraw-Hill Inc. (1995)