

# Analyse vidéo : avancées et opportunités au cœur de la révolution numérique

Patrick Pérez  
IRISA/INRIA-Rennes

Irisatech  
26 juin 2007

# Contexte

---

- **Convergence** : internet, PC multimedia, téléphones, PDA, lecteurs multimédia portables, TV
- **Puissance de calcul** : RAM, processeurs double-cœur, GPU, grilles
- **Bases de vidéos** : P2P, YouTube, DailyMotion, SoapBox, INA, BBC
- **Nouveaux matériels**
  - set-top boxes
  - HD-DVD et Blu-Ray
  - caméscopes numériques avec DVD
  - appareils photos numériques avec vidéo

# Domaines

---

- Annotation, recherche, préparation, diffusion et protection de contenus vidéos (et multimédia)
  - descripteurs compacts et recherche rapide
  - recherche d'instance ou de classe d'objets, de scène, d'action; recherche de copies vidéos
  - découpage en plan, résumé automatique, adaptation des contenus aux périphériques, création et insertion de méta-données ou de marques
- **Graphique**: AR/VR, capture de mouvement (jeux, films et avatars)
- **Post-production** (cinéma et TV): AR, édition, colorisation, enlevage d'objets, flous sélectifs, ralentis, restauration, stabilisation
- **Vidéos sportives**: analyse de gestes et d'actions, AR/VR, statistiques de joueurs, annotation de vidéos

# Avancées récentes

---

- **Appariement** (extraction, description et mise en correspondance de points d'intérêt, etc.)
- **Machine learning** et théorie de l'apprentissage (*belief propagation*, SVM, méthodes à noyaux, *boosting*, etc.)
- **Méthodes probabilistes**, paramétriques ou non (modèles graphiques, filtrage stochastique, etc.)
- **Méthodes basées exemples** (*exemplars*, *patches*, etc.)
- **Optimisation combinatoire** sur graphes (*min-cut/max-flow*, etc.)
- **EDPs pour l'image** (*level sets*, etc.)
- **Représentations parcimonieuses** (ondelettes, atomes temps-fréquences, etc.)

# Tâches génériques d'analyse

---

- Analyse d'images fixes
  - segmentation (partitionnement)
  - extraction et appariements de descripteurs locaux
  - détection, reconnaissance, segmentation d'un objet donné ou issu d'une classe donnée d'objets
- Analyse du mouvement
  - estimation de champs de mouvements (« flots optiques ») instantanés
  - estimation paramétrique du mouvement dominant et détection des zones de mouvement résiduel
  - segmentation au sens du mouvement
  - extraction et appariement de descripteurs dynamiques locaux
  - suivi de régions 2D ou de modèles 3D d'objets

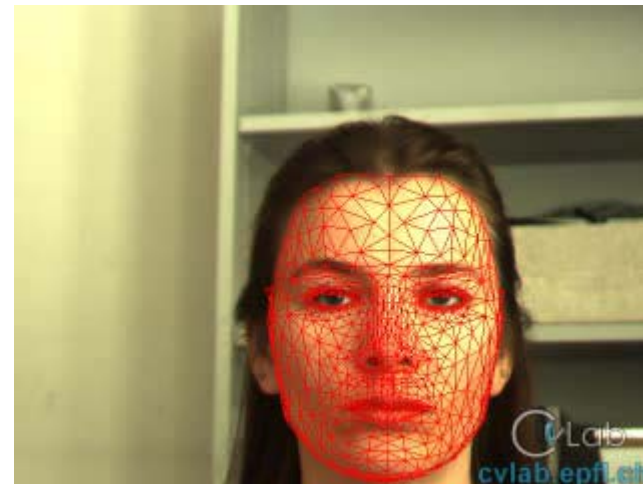
# Estimation/suivi de pose en VR/AR

---

- Modèles rigides 3D



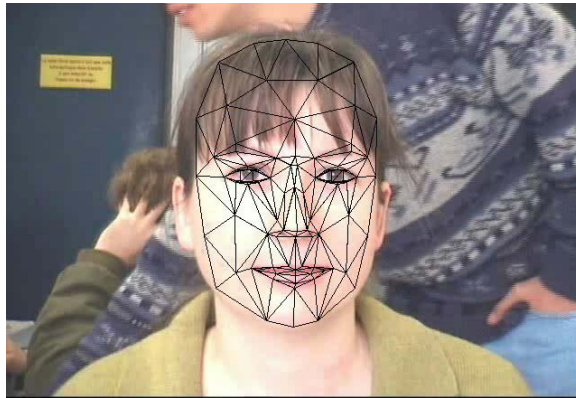
[Comport *et al.*, 2003]



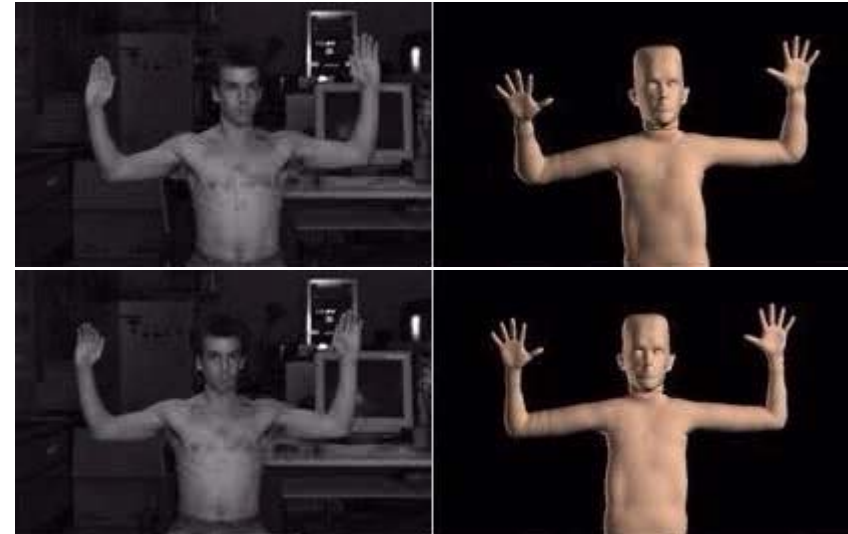
[Vacchetti et al., 2003]

# Estimation/suivi de pose en VR/AR

- Modèles déformables et/ou articulés



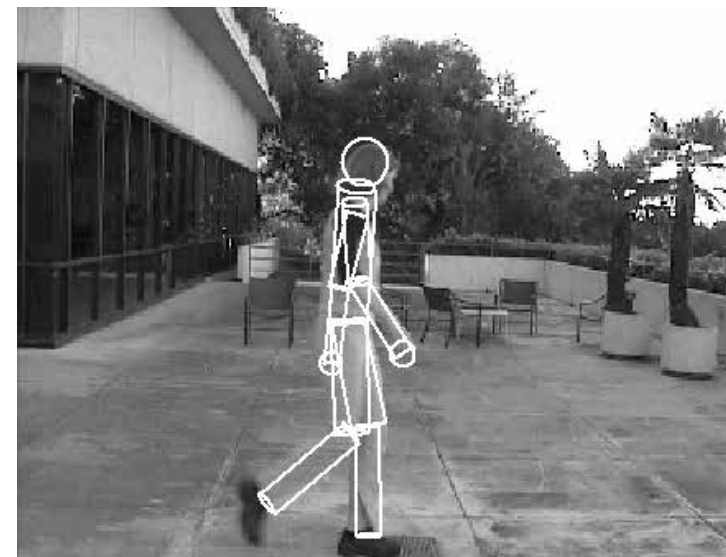
[Davoine *et al.*, 2004]



[Plaenkers and Fua, 2003]



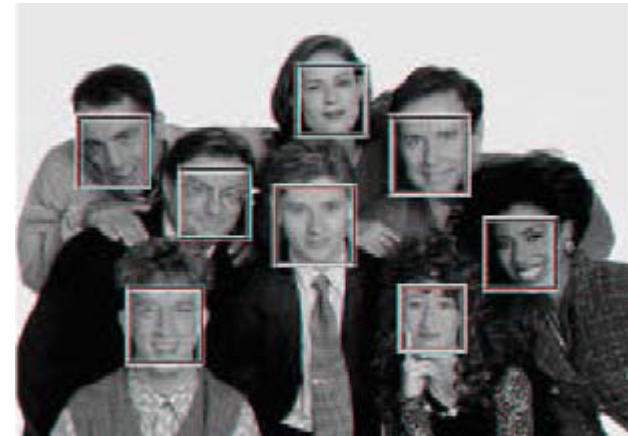
[Sudderth *et al.*, 2004]



[Sidenbladh *et al.*, 2000]

# Détection et suivi de classes d'objets

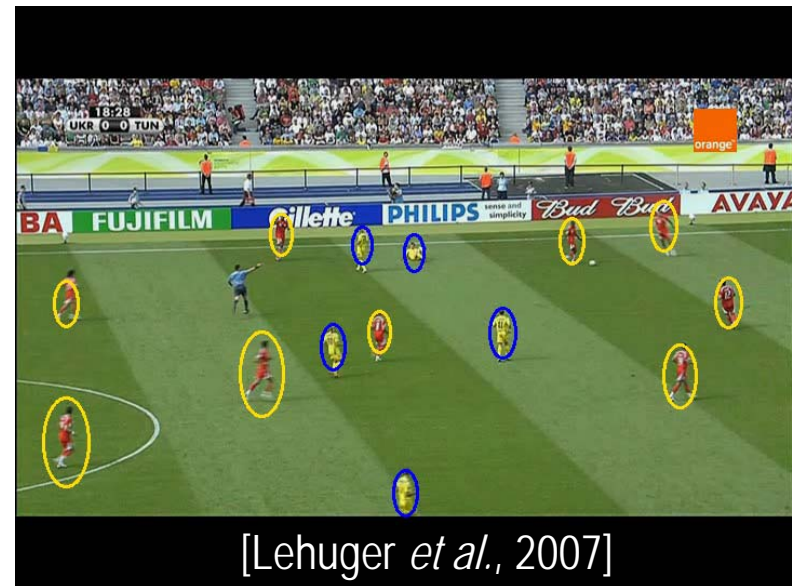
- **Détection** : classifieur SVM ou *boosting* appris hors-ligne
- **Suivi** : recherche déterministe de proche en proche ou filtrage particulière



[Viola and Jones, 2001]



[Okuma *et al.*, 2004]



[Lehuger *et al.*, 2007]



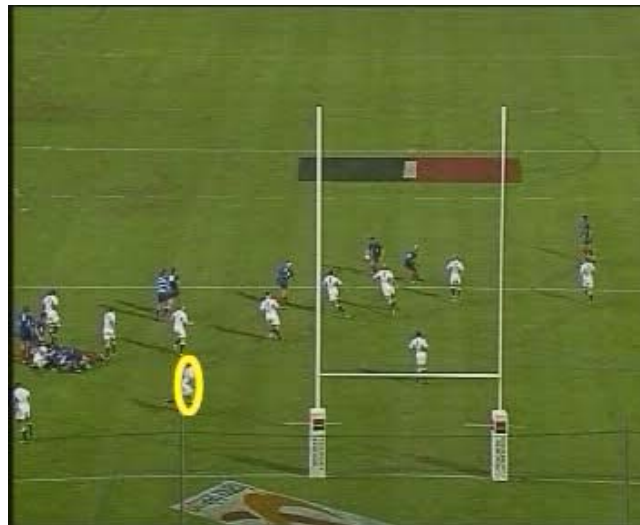
# Suivi d'« objets » arbitraires

- Modélisation de l'apparence : *templates*, histogrammes de couleur, formes simples
- Suivi : recherche déterministe ou filtrage particulière

[Comaniciu et al., 2001]



[Pérez et al., 2002]



[Gengembre and Pérez, 2006]

[Badrinarayanan *et al.*, 2007]



# Recherche dans une vidéo

- Film prétraité : points d'intérêts extraits, caractérisés, tabulés
- Requête (étendue par suivi) : un "objet"
- Résultat : autres instances de l'objet dans le film (ou d'autres vidéos)

← suivi →



"VideoGoogle" at Oxford Visual Geometry Group [Sivic *et al.*, 2004]

# Résumé de vidéo

- Résumés statiques : extraction d'une ou plusieurs images clés par plan

[Vermaak et al. 2002]



- Résumés vidéos spécifiques :  
exemple du foot (analyse du son, de la couleur, des mouvements de caméra)

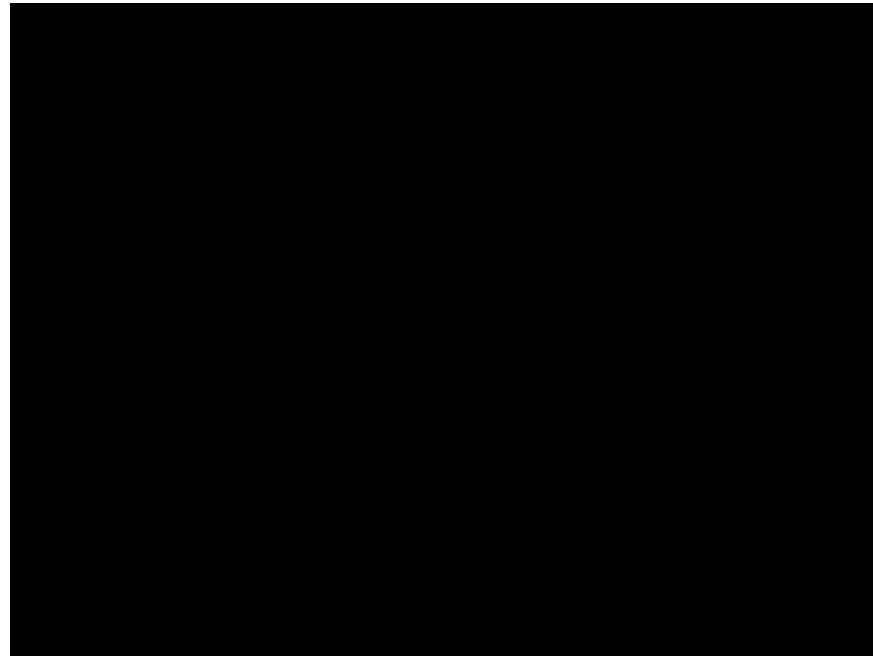
[N. Peyrard, P. Bouthemy 2005]



# Reconnaissance d'action

---

- **Apprentissage supervisé** : annotation de « volumes » vidéos pour un ou plusieurs types d'action; primitives mouvements (histogrammes locaux de flots optiques); apprentissage par *boosting*
- **Détection hors-ligne**



[Laptev and Pérez, 2007]

# Stabilisation

---

- **Problème** : estimation et compensation du mouvement dominant
- **Outils** : régression paramétrique robuste
- **Applications** : restauration, amélioration en vidéo professionnelle ou non



[Matsushita *et al.* 2005]

# Enlevage d'objets

- **Outils** : estimation du mouvement, suivi, synthèse de texture par l'exemple
- **Applications** : restauration, corrections, effets spéciaux



[Wexler *et al.* 2007]

# Cutout

---

- **Segmentation** d'un objet au cours du temps
- **Outils** : estimation du mouvement ; segmentation par la couleur à l'aide de *min-cut/max-flow* ; extraction de *matte*
- **Application** : sélection pour édition (e.g., rendu non-réaliste), masquage ou *compositing*



[Wang *et al.*, 2005]

# Analyse de gestes sportifs

---

- Vitesses et trajectoires par suivi avec *a priori*



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

- Alignement de séquences



DartFish's SimulCam



# Défis

---

- Très grands volumes (archives vidéos immenses, images HD)
  - analyse *off-line* si appropriée
  - optimisation des traitements *on-line* (représentation compacte, recherches approchées, etc.)
- Multimodalité, multidisciplinarité
  - combiner analyse d'images avec analyse du son (voix en particulier) et du texte (sous-titres, génériques, scripts, incrustations)
  - combiner analyse et synthèse
  - associer recherche et R&D
- Robustification
  - complexité et variabilité des contenus défient les outils actuels d'analyse d'images fixes et de vidéos
  - occlusions : problème pour l'analyse du mouvement et le suivi
- Nouveaux usages : Méta-données basées suivi ? Types de requêtes ? Nouveaux outils semi-automatiques d'édition et post-production ? Professionnels vs. Amateurs?