Apprentissage continu de représentations visuelles

ENSIMAG
2023-2024

Karteek Alahari & Diane Larlus

https://project.inria.fr/bigvisdata/

Informations : Rappel

- Évaluation
  - Examen final écrit
  - Quizz sur des articles de recherche
  - https://project.inria.fr/bigvisdata/grading

- Points bonus
  - Presentation de papier
  - Voir la liste : https://project.inria.fr/bigvisdata/presentations
  - Votre choix par email

Aperçu du cours

Cours 1 : introduction

- Définition ‘Big Data’, applications, la chaine, lien avec la recherche d’information par le contenu
- List des thèmes abordés: apprentissage supervisé, auto-supervisé, adaptation de domaine, apprentissage continu, problèmes en vidéos
- Définition de l’apprentissage supervisé et de ses étapes majeures: collection des données, choix d’une représentation, choix d’un modèle, apprentissage, sélection d’un modèle
- Collection des données: difficulté et ambiguïté de l’annotation d’images, exemples de grandes bases d’images standard
- Décomposition de l’erreur
- Représentation des données: approches traditionnelles: représentations puis apprentissage, *mid-level representations*, apprentissage de bout-en-bout
- Définition de l’apprentissage profond

Aperçu du cours

Cours 1 : introduction

- Définition ‘Big Data’, applications, la chaine, lien avec la recherche d’information par le contenu
- List des thèmes abordés: apprentissage supervisé, auto-supervisé, adaptation de domaine, apprentissage continu, problèmes en vidéos
- Définition de l’apprentissage supervisé et de ses étapes majeures: collection des données, choix d’une représentation, choix d’un modèle, apprentissage, sélection d’un modèle
- Collection des données: difficulté et ambiguïté de l’annotation d’images, exemples de grandes bases d’images standard
- Décomposition de l’erreur
- Représentation des données: approches traditionnelles: représentations puis apprentissage, *mid-level representations*, apprentissage de bout-en-bout
- Définition de l’apprentissage profond

Cours 2 : introduction + adaptation de domaine

Quelques notions utiles

Comment créer des représentations visuelles ?

- Entrainer un modèle sur de grandes quantités de données annotées
- Utiliser ce modèle pour produire une représentation vectorielle pour chaque image proposée en entrée du modèle

Comment réutiliser des représentations visuelles ?

- Utiliser des représentations visuelles directement, optionnellement apprendre un modèle de décision par-dessus
- Utiliser le modèle précédent comme point de départ pour l’apprentissage et l’ajuster (*fine-tuning*) pour la tâche cible
- Appliquer une méthode d’adaptation, par exemple: adaptation de domaine

Comment créer des représentations visuelles quand on n'a pas d’annotations ?

- Introduction à l’apprentissage auto-supervisée

Dans ce cours

- Apprentissage supervisé (*supervised learning*)
  - Variantes, ex. Semi-supervisé (*semi-supervised*)
- Apprentissage auto-supervisé (*self-supervised*)
- L'adaptation de domaine (*domain adaptation*)
- Apprentissage continu (*continual learning*)
- **Problèmes en vidéos**

Reconnaissance d’actions

Action classification in videos

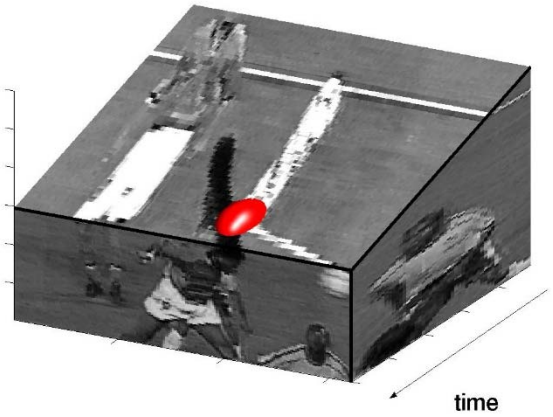
- Space-time interest points [Laptev, IJCV'05]
- Dense trajectories [Wang and Schmid, ICCV'13]
- Video-level CNN features

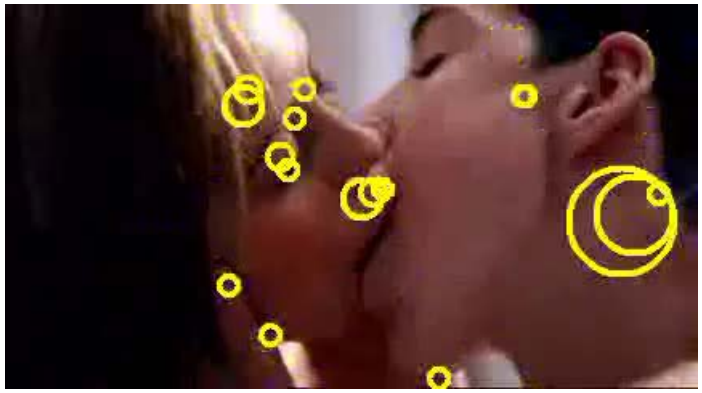
Space-time interest points (STIP)

- Space-time corner detector
  [Laptev, IJCV 2005]

$$H = \det(\mu) + k\,\mathrm{tr}^3(\mu)$$

$$\mu = \begin{pmatrix} I_x I_x & I_x I_y & I_x I_t \\ I_x I_y & I_y I_y & I_y I_t \\ I_x I_t & I_y I_t & I_t I_t \end{pmatrix} * g(\cdot;\, \sigma, \tau)$$

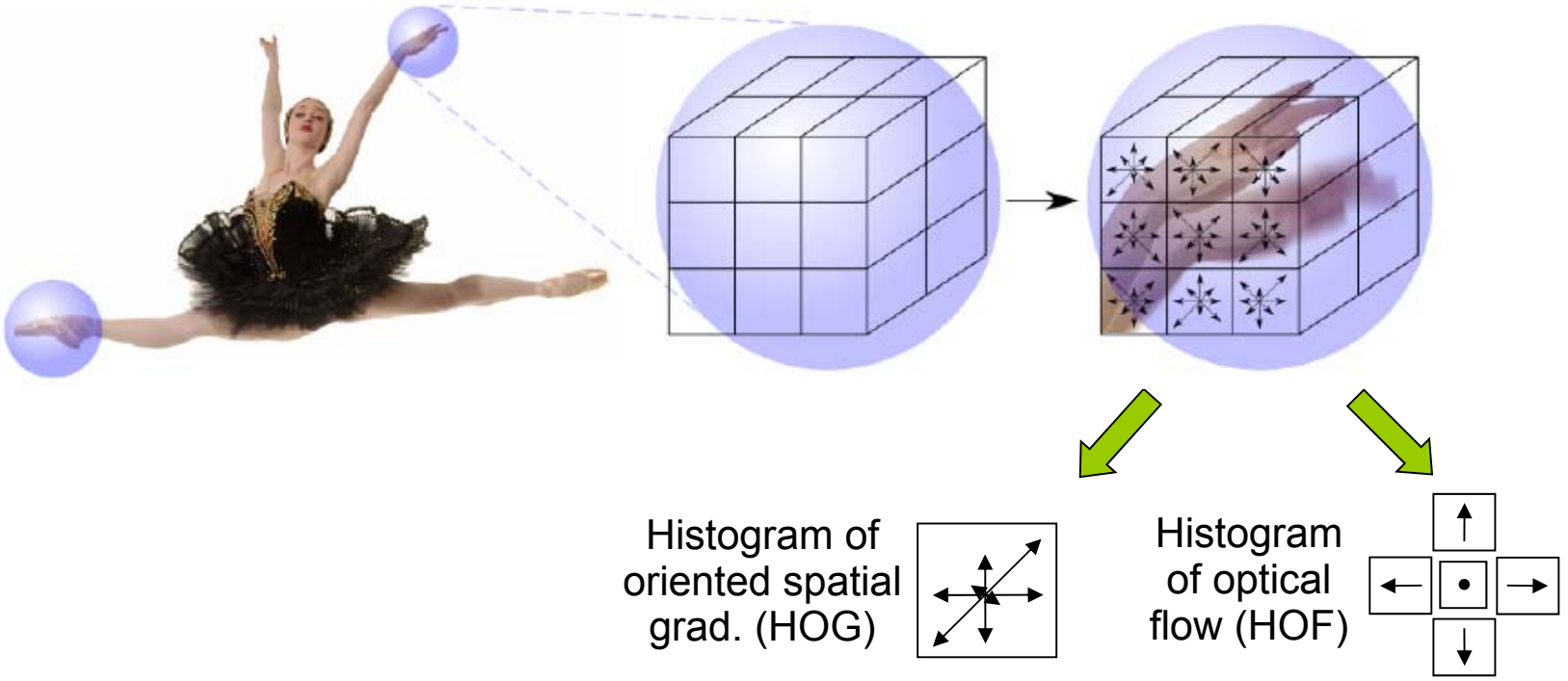




informatics mathematics

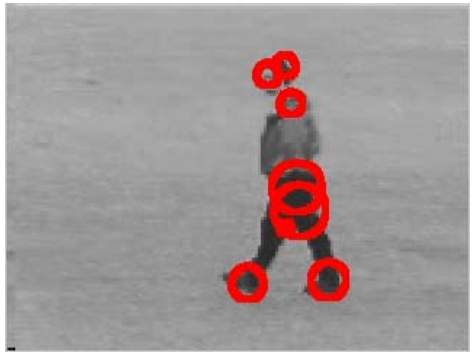
STIP descriptors

Space-time interest points

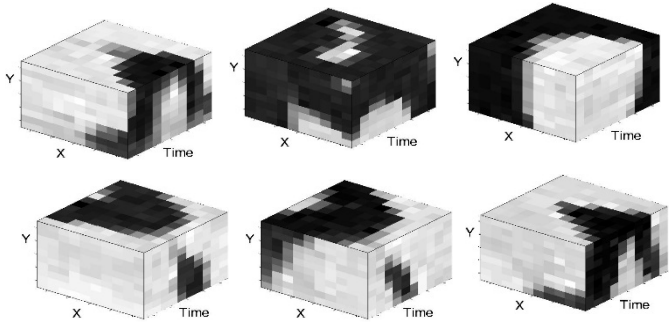


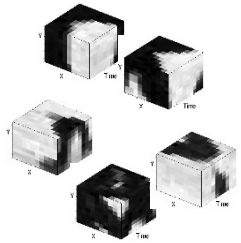
Action classification

- Bag of space-time features + SVM [Schuldt'04, Niebles'06, Zhang'07]



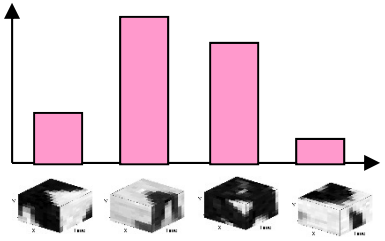
Collection of space-time patches





HOG & HOF patch descriptors

Histogram of visual words

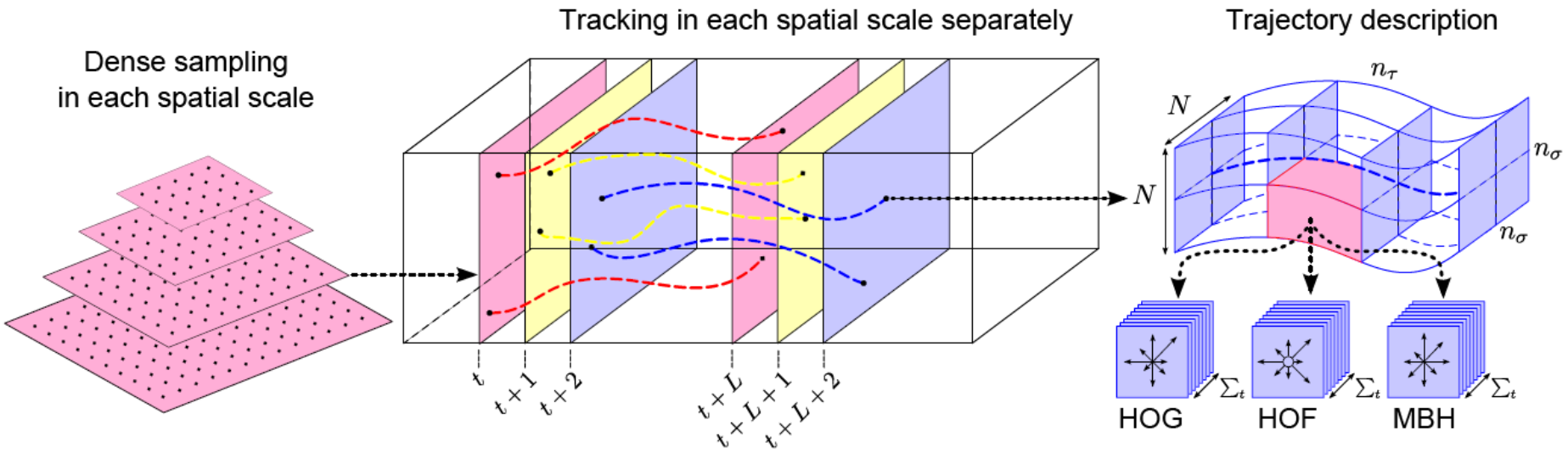


SVM Classifier

informatics mathematics

State of the art for video description

- Dense trajectories [Wang et al., IJCV'13] and Fisher vector encoding [Perronnin et al. ECCV'10]

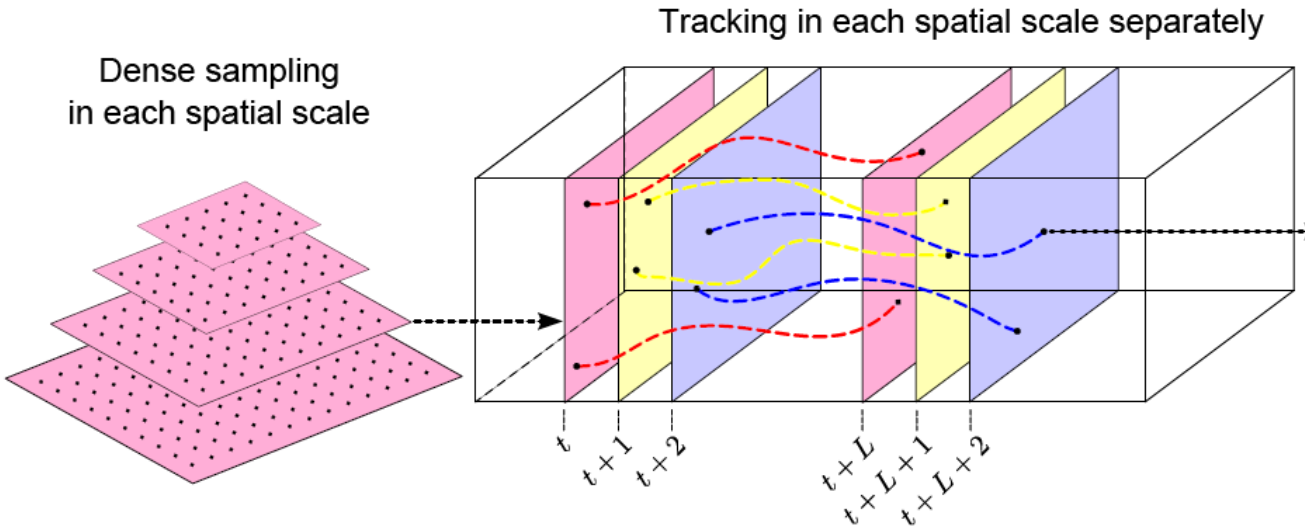


- Orderless representation

informatics mathematics

Dense trajectories [Wang et al., IJCV'13]

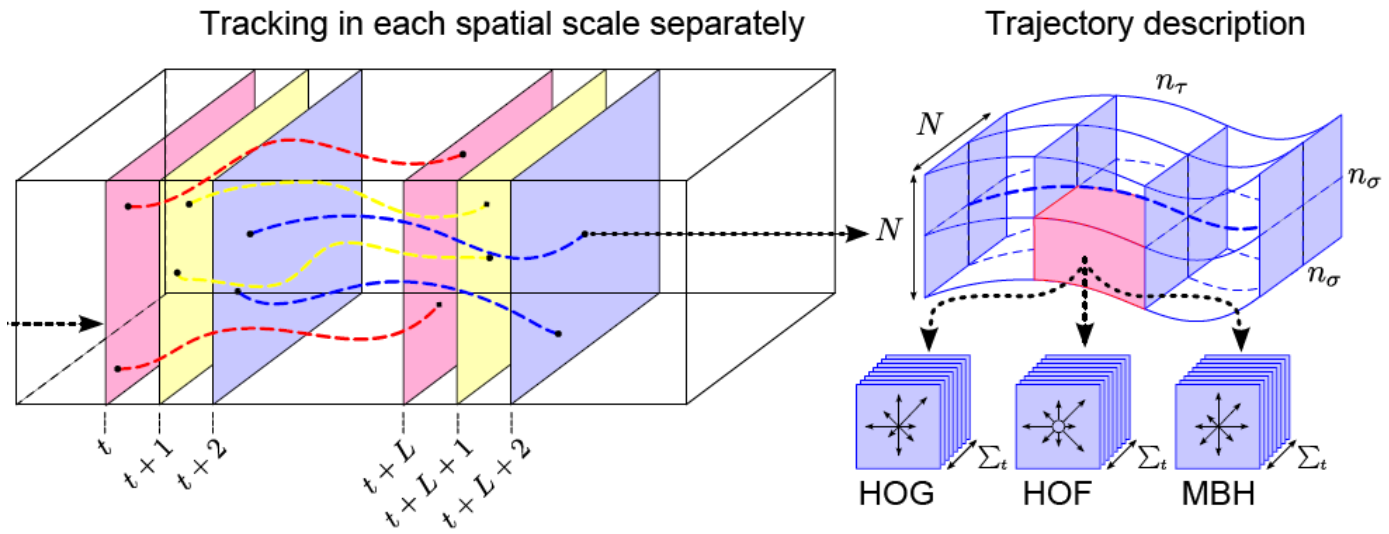
- Dense sampling at several scales
- Feature tracking based on optical flow for several scales
- Length 15 frames, to avoid drift



informatics mathematics

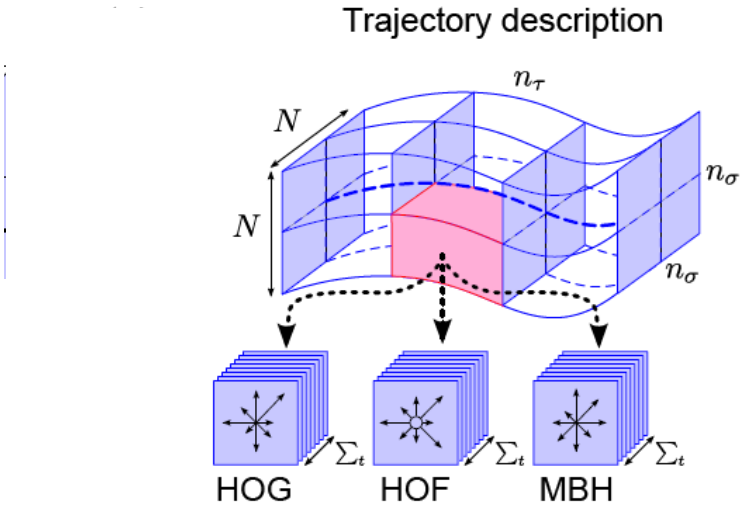
Descriptors for dense trajectory

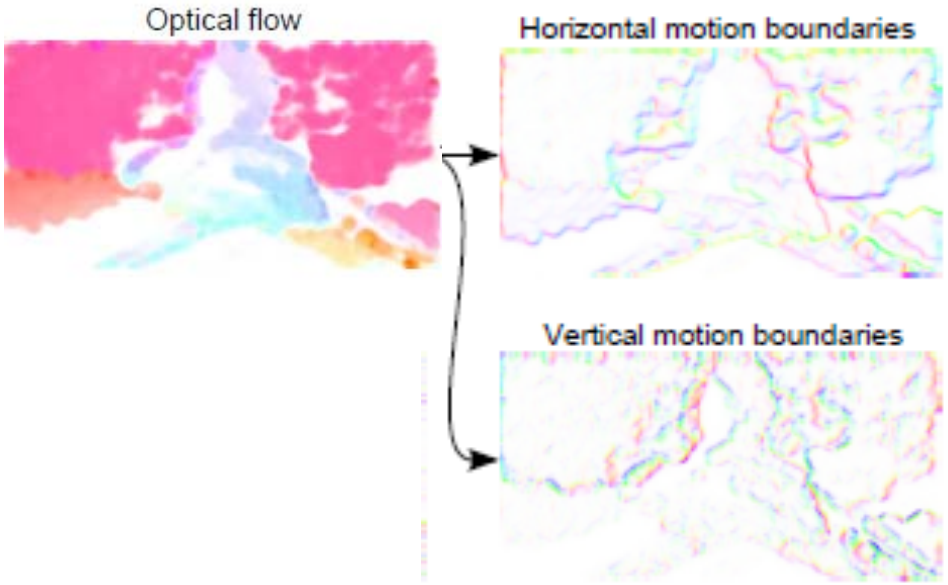
- Histogram of gradients (HOG: 2x2x3x8)
- Histogram of optical flow (HOF: 2x2x3x9)



Descriptors for dense trajectory

- Motion-boundary histogram (MBHx + MBHy: 2x2x3x8)
  - spatial derivatives are calculated separately for optical flow in x and y, quantized into a histogram
  - captures relative dynamics of different regions
  - suppresses constant motions





informatics mathematics

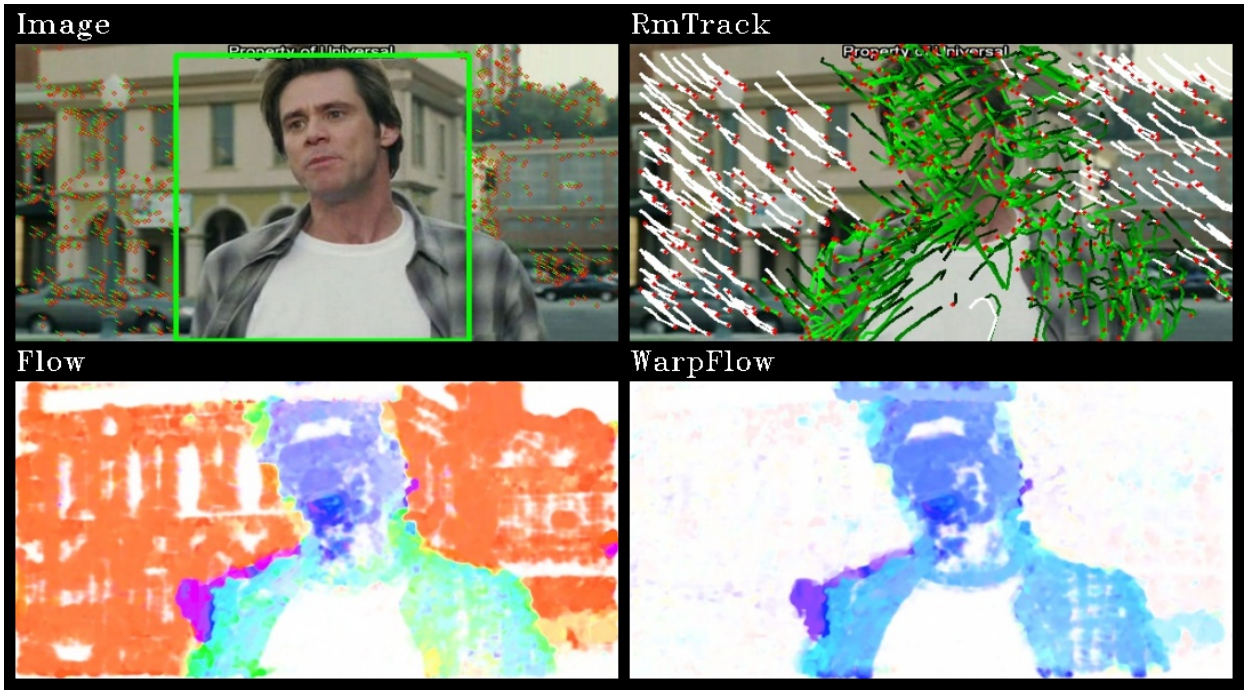
Dense trajectories

- Advantages:
  - Captures the intrinsic dynamic structures in videos
  - MBH is robust to certain camera motion

- Disadvantages:
  - Generates irrelevant trajectories in background due to camera motion
  - Motion descriptors are modified by camera motion, e.g., HOF, MBH

Improved dense trajectories

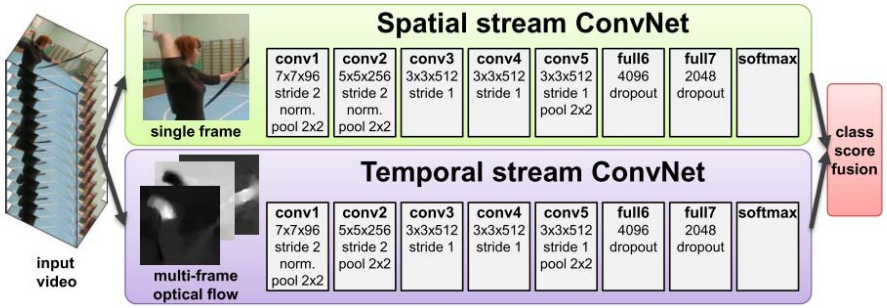
- Improve dense trajectories by explicit camera motion estimation
- Detect humans to remove outlier matches for homography estimation
- Stabilize optical flow to eliminate camera motion



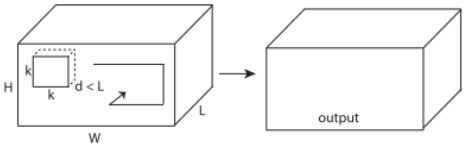
[Wang and Schmid. Action recognition with improved trajectories. ICCV'13]

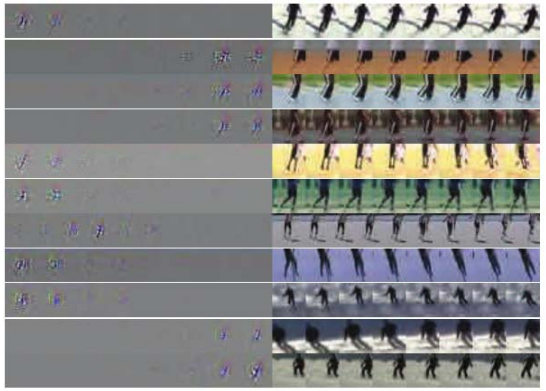
Recent CNN methods

Two-Stream Convolutional Networks for Action Recognition in Videos [Simonyan and Zisserman NIPS14]



Learning Spatiotemporal Features with 3D Convolutional Networks [Tran et al. ICCV15]





Quo vadis action recognition? A new model and the Kinetics dataset [Carreira et al. CVPR17]