Could early visual processes be sufficient to label motion?

08/2005

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

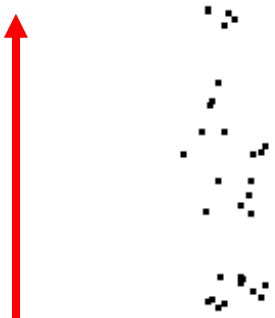
Pierre Kornprobst, Thierry Vieville, Ivan Dimov
INRIA, Odyssée Team
France

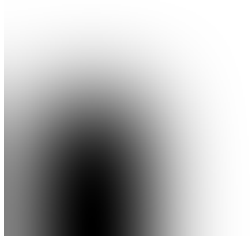
Overview

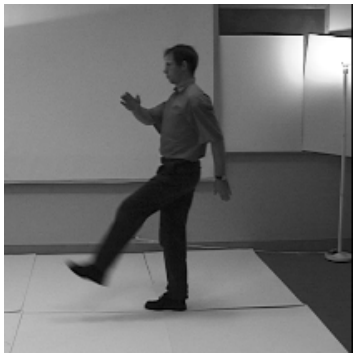
- Our concern is biological motion recognition

- Early visual processes refers to rank order coding schemes

- Given videos, our aim is to find an automatic approach which discriminates between different categories







Plan

Introduction

State of the Art

System Overview

Results and Discussion

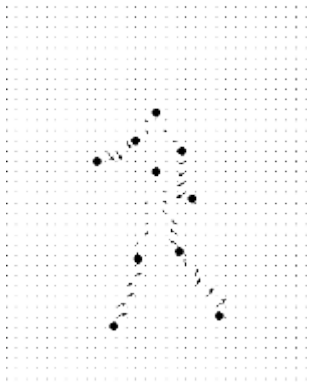
Future work

INRIA

State of the Art in Psychophysics

- Input are synthesized moving point sequences or points trajectories manually selected.

- Two pathways involved:



and/or

The **motion** pathway the **form** pathway

- References include Giese, Casiles [2005], Beintema, Lappe [2002]

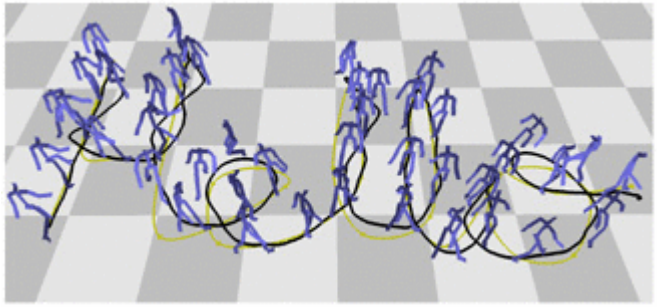
Motion information is most important, only coarse form information needed

State of the Art in Computer Graphics

Motion capture system
VICON*PEAK*

- The input are usually points given by motion capture systems

- These systems allow a precise analysis of body motions.

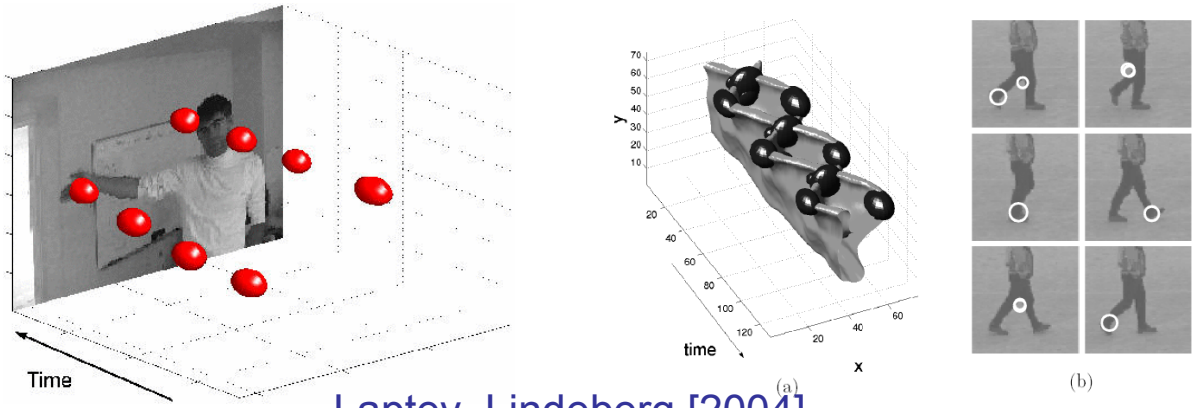


Motion graphs
Kovar, Gleicher, Pighin [2002]

- Points models can also be used to generate new smooth motions motion capture databases, and animate some avatars.

State of the Art in Computer Vision

- The input is a monocular or stereo sequence of images.
- There exists a wide litterature with a variety of approaches based on
  - Human 3D-model recovery,
  - Motion body parts tracking,
  - Motion periodicity analysis, etc.
- In some recent work, Laptev and Lindeberg [2004] proposed a method to use some spatio-temporal features for action recognition



Laptev, Lindeberg [2004]

Plan

Introduction

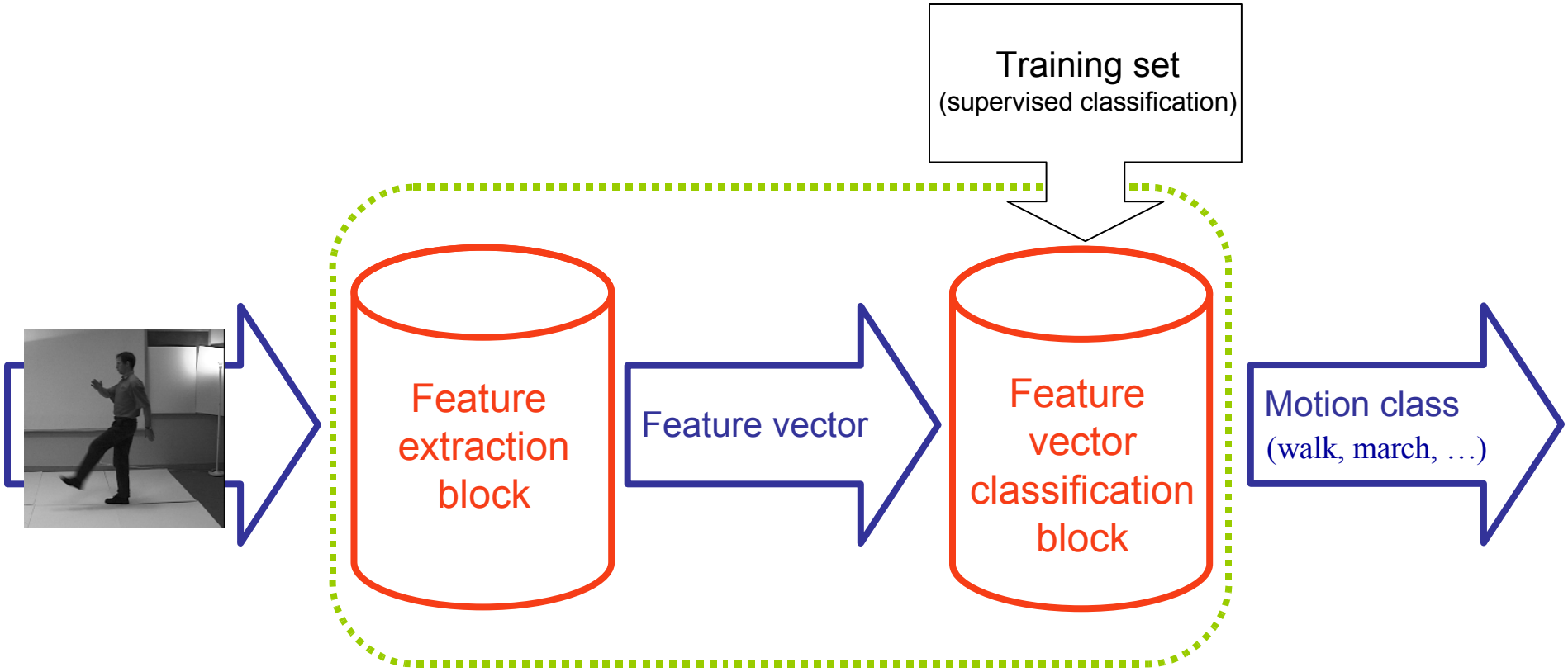
State of the Art

System Overview

Results and Discussion

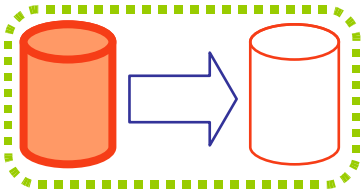
Future work

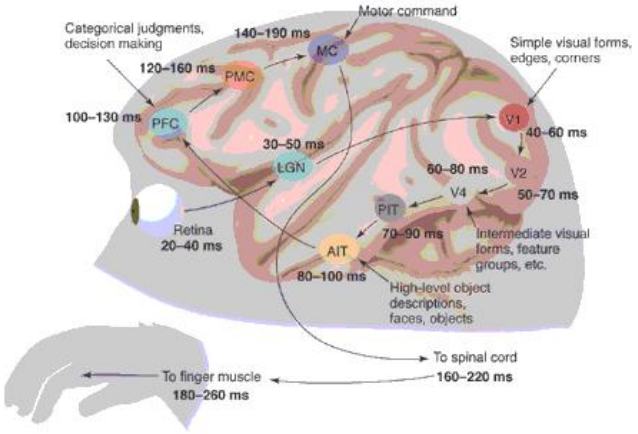
INRIA



Retina V1, V2, V4, V5 IT

Feature Extraction Block





Thorpe etal [1996, 2002]

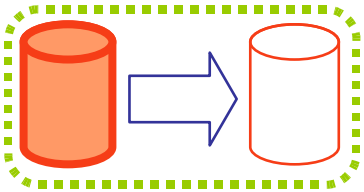
- The fact: Primates can categorize complex visual stimuli very quickly, with reaction times that can be as short as 150 ms.

- From standart coding (spike count, mean interspike interval, mean firing rates) to rank order coding

The first ganglion cells to emit a spike are given a maximal weight

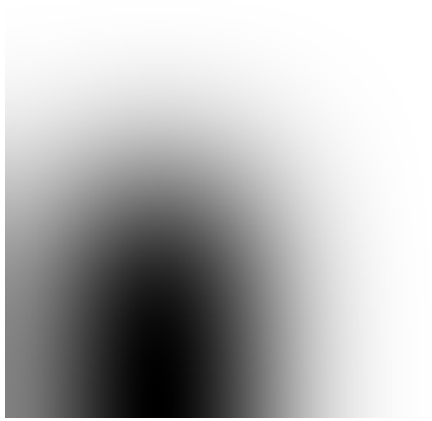
- The rank codes are optimal for fast information transmission

Feature Extraction Block



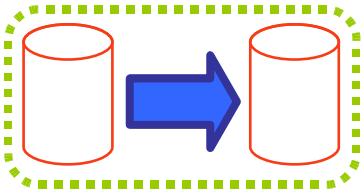
- Simulation of rank order coding can be implemented using matching pursuit-like algorithm

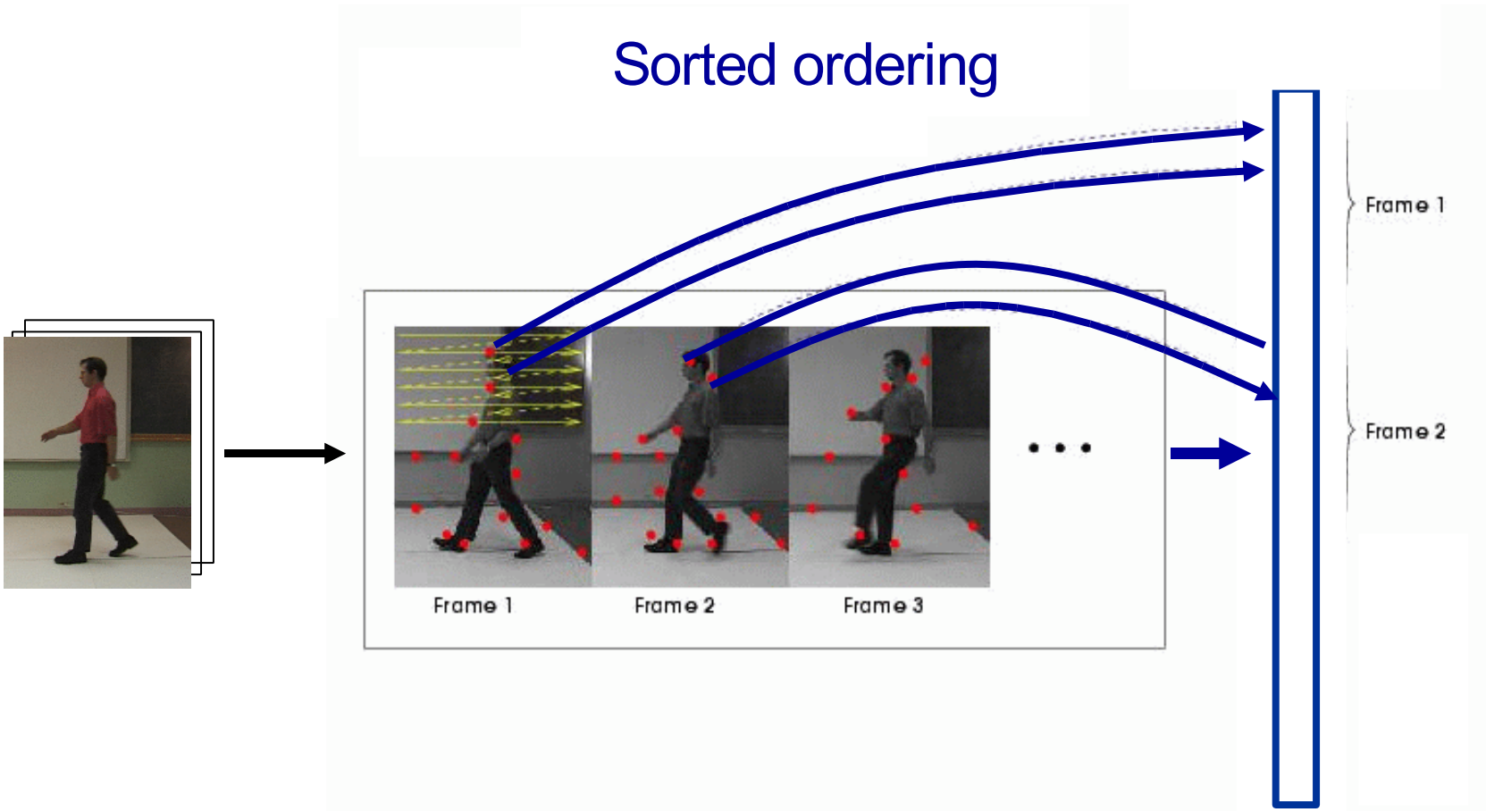
- Extension to video coding: Several possibilities exist. The simplest one in this work.



- References: Mallat [1993], Perrinet, Samuelides and Thorpe [2004], Olsehausen [2003]).

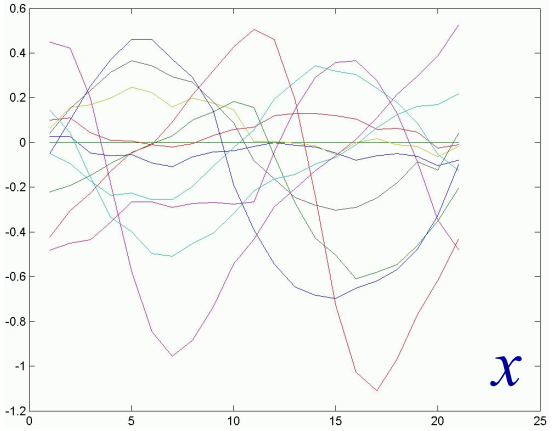
The Feature Vector

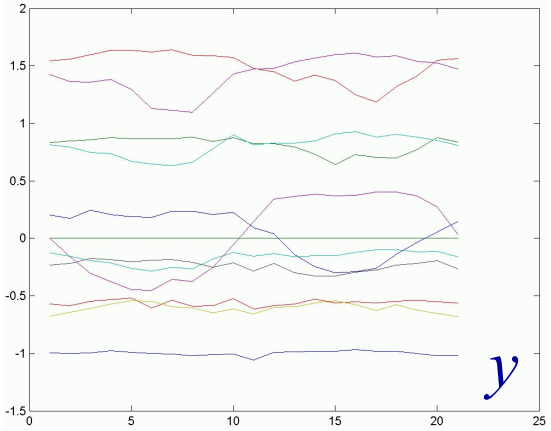




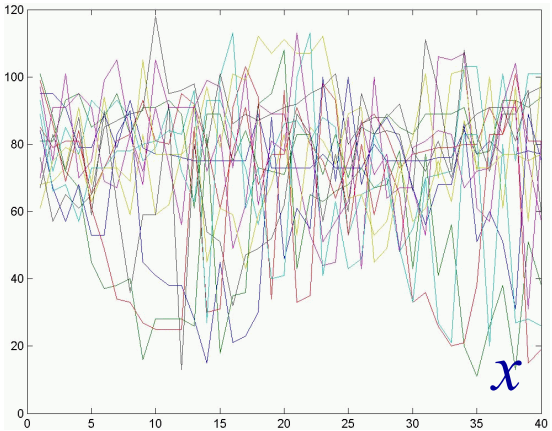
Example of trajectories

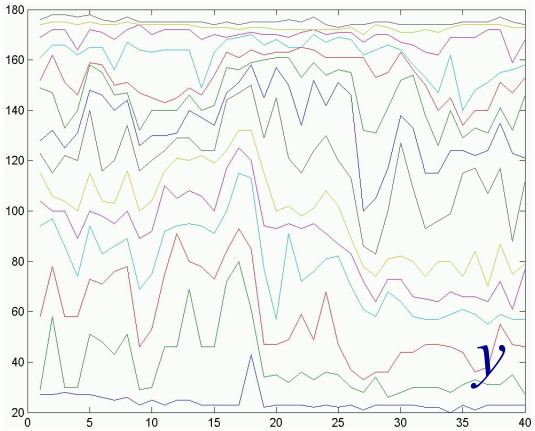
- Trajectories of manually marked joint positions



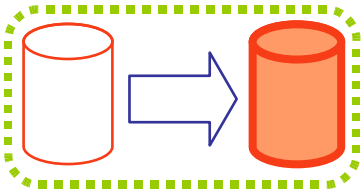


- Trajectories generated by spikes positions in the feature vector





Feature Vector Classification Block



• RAW classifiers are simple nearest-neighbor mechanisms: the category of the closest prototype defines the category.

• SVM classifiers are optimal mechanims of classification given a set of prototypes (supervised learning)

• It is known that SVM classifiers
  - can be implemented using biologically plausible Hebbian mechanisms
  - can be interpreted as (optimized) nearest-neighbor mechanisms.

Vieville, Crahay [2004]

Plan

Introduction

State of the Art

System Overview

Results and Discussion

Future work

INRIA

40 biological motion image sequences
2 classes of motions

(Courtesy M. Giese)

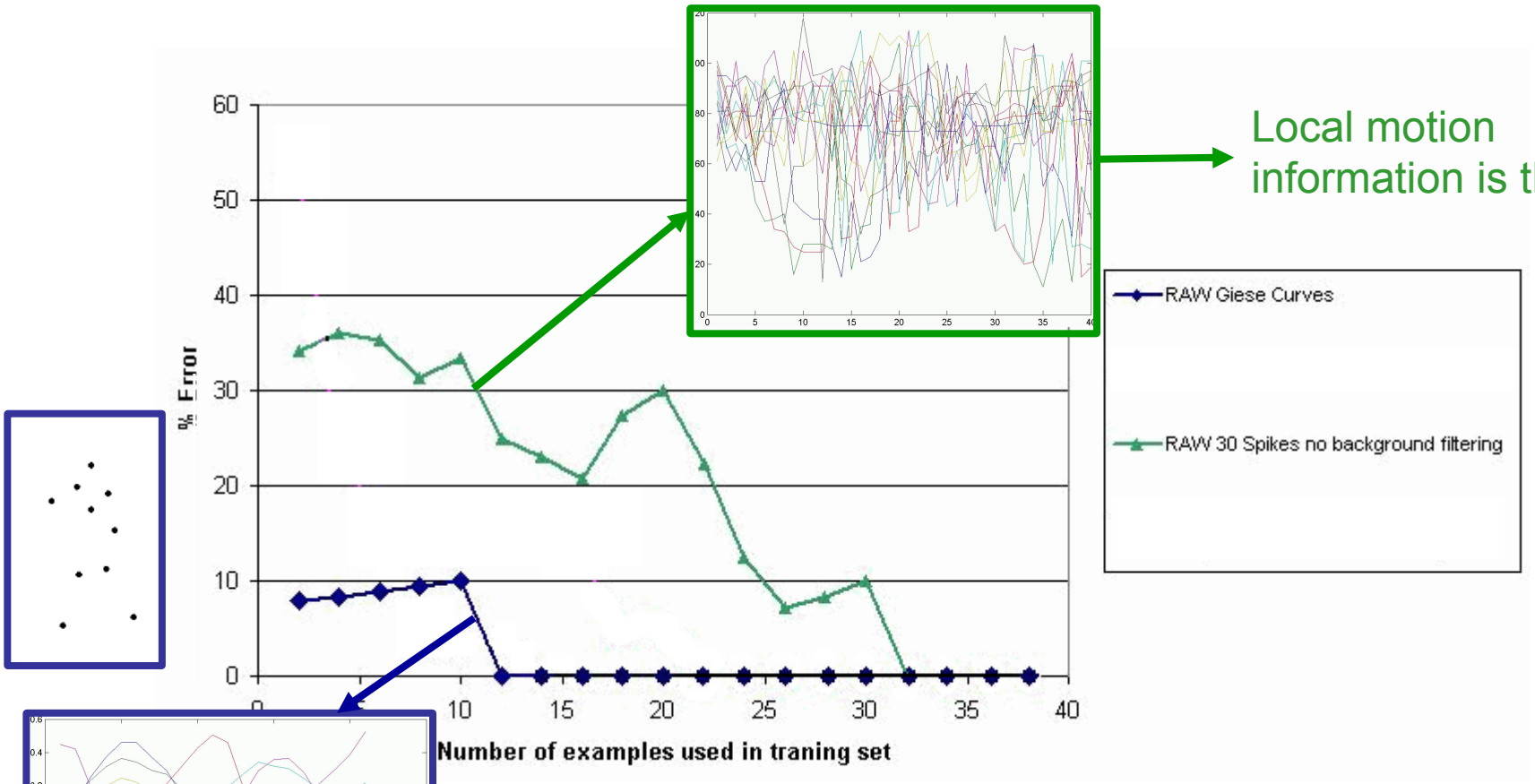
March

Walk

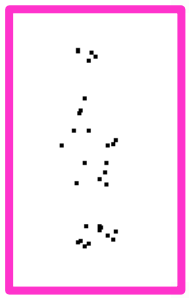
- Learning phase with one randomly chosen feature vector from each class and randomly repeatedly incrementing the data sample in the walking and marching classes

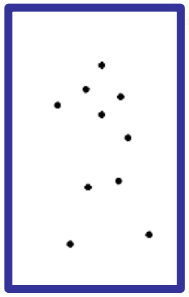
- When the training is completed the remaining data samples are used as testing set to quantify the classifier's error rate.

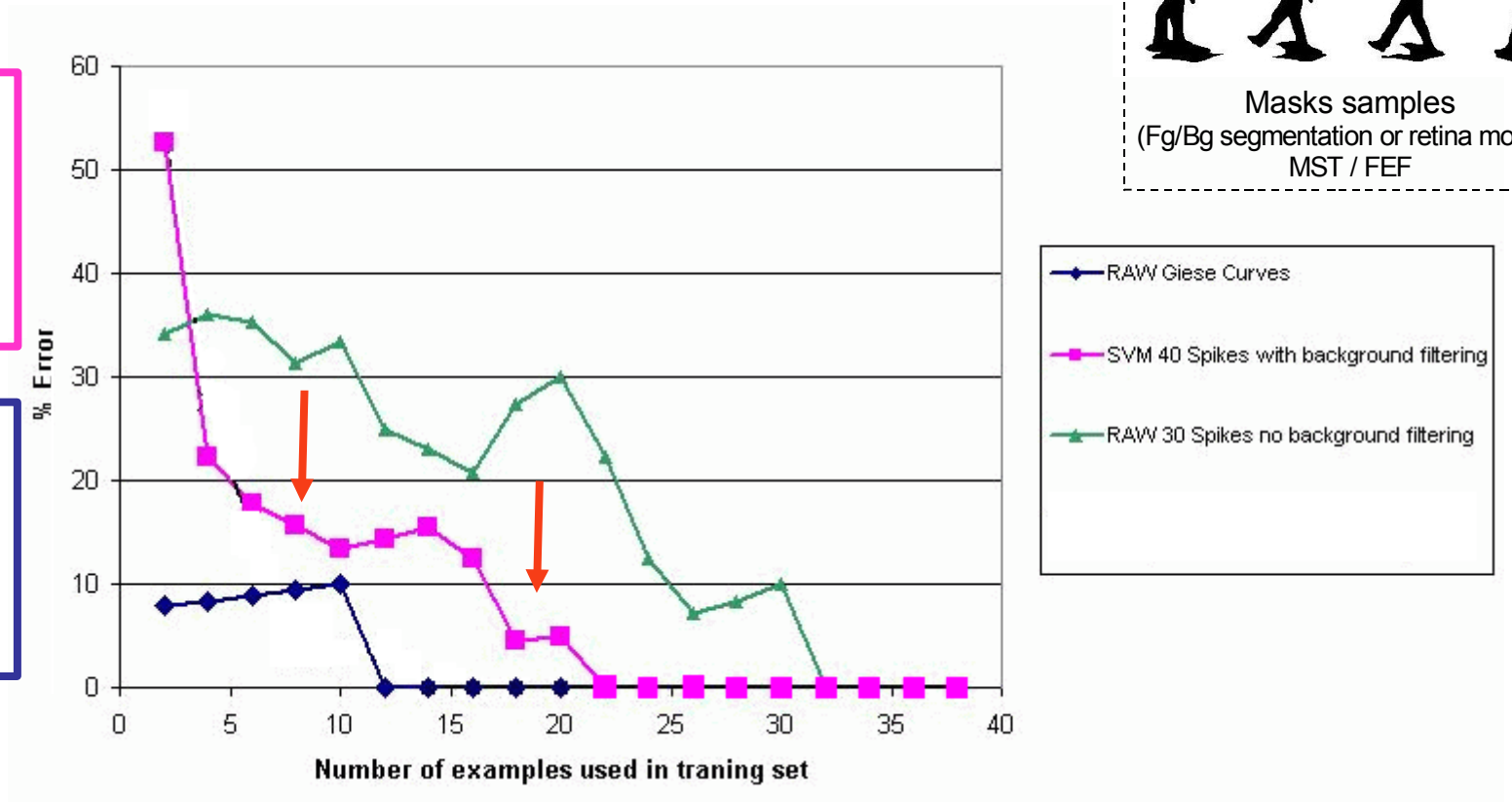
Joints Trajectories vs. Spike Coding



Filtering Spikes from the Background Improves Classification

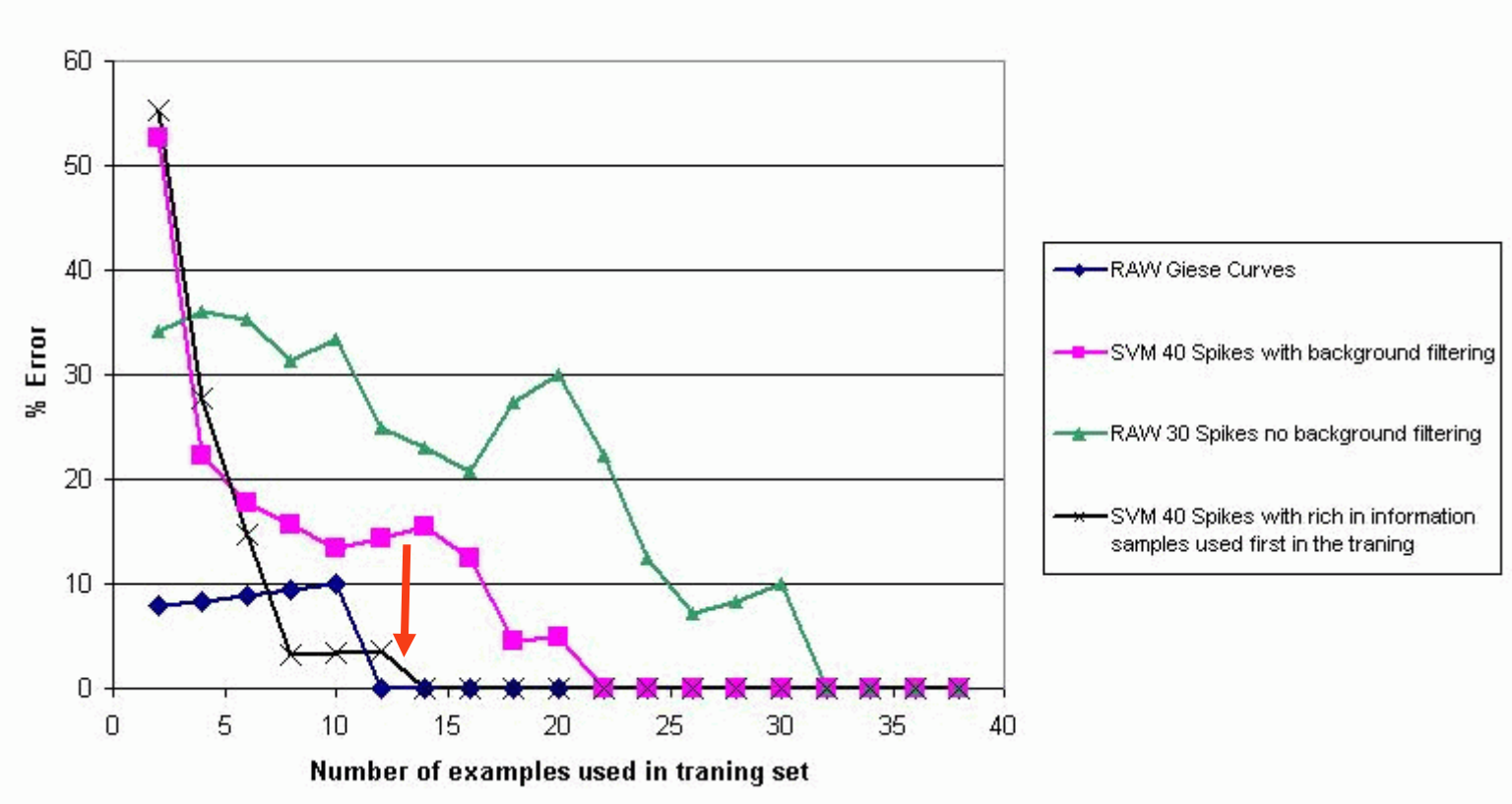






Masks samples
(Fg/Bg segmentation or retina motion)
MST / FEF

Some Sequences Bring More Informations to the Training Phase



Plan

Goals

State of the Art

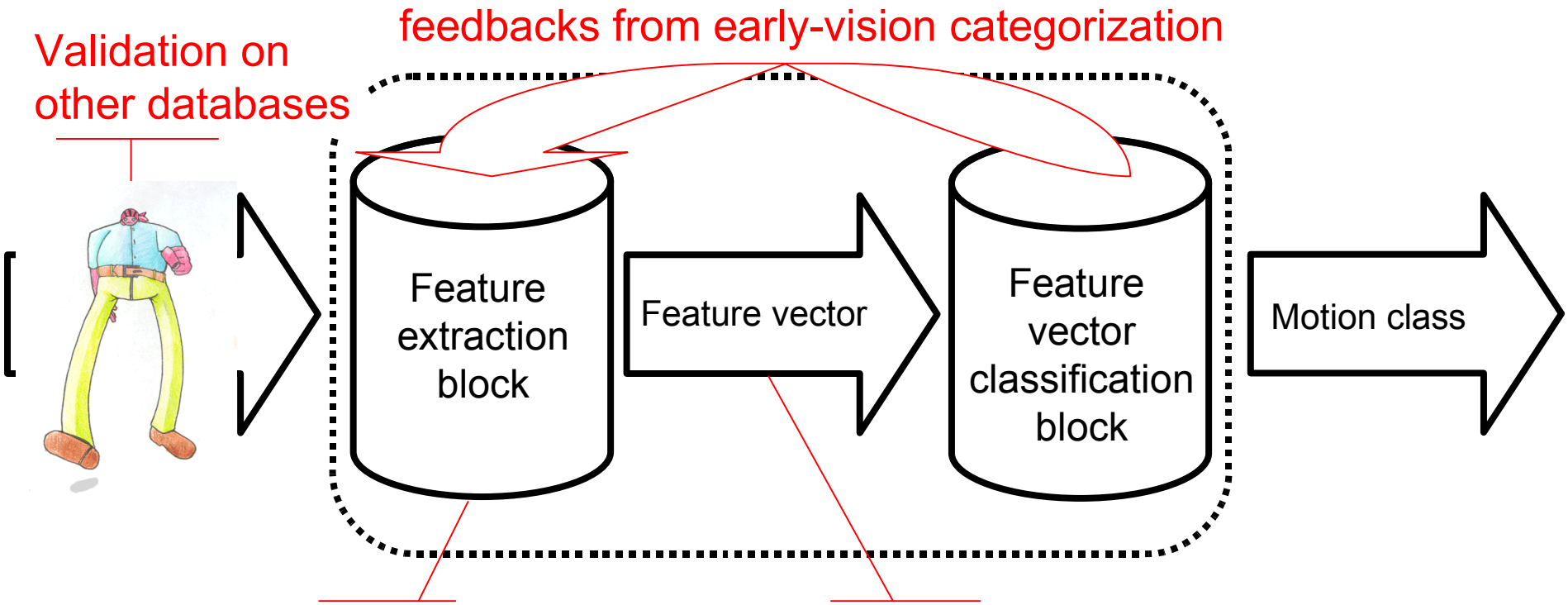
System Overview

Results and Discussion

Future work

INRIA

Could early visual processes be sufficient to label motion? **Yes**, but…



From categorization to segmentation

- Feedback from fast-brain object categorization likely help segmentation (here object-background segmentation)

  Friston [2002], Bullier [2001]

- In fact there is an interaction between both mechanisms

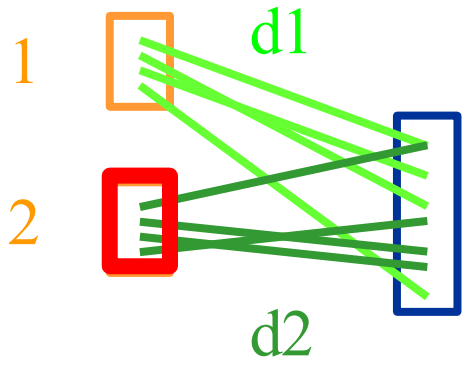
  - It is assumed that during the learning phase the spikes related to the object are likely identified

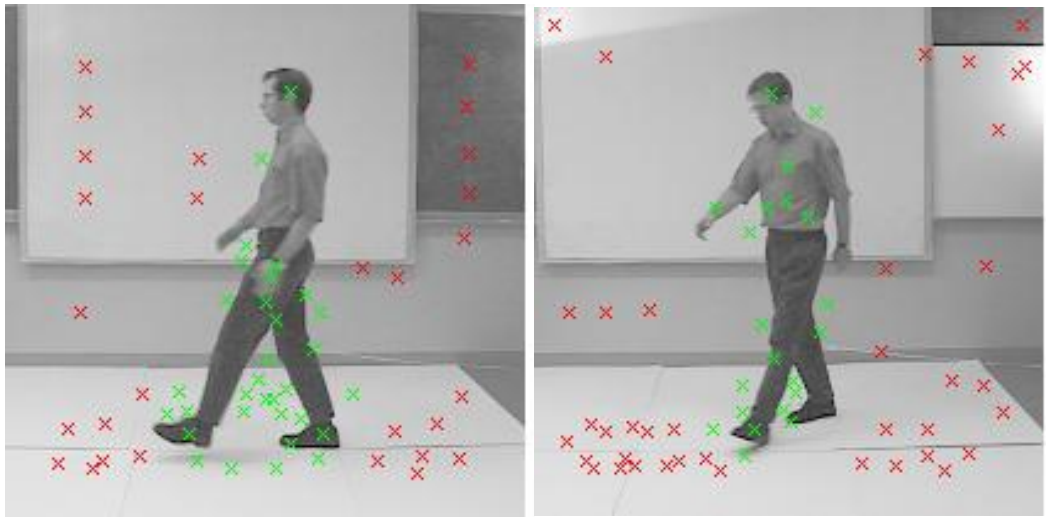
  - During the categorization phase, spikes from both background and object are mixed : the better the segmentation the better the categorization

- Segmentation can be derived considering spikes close to the related prototype's spikes

From categorization to segmentation

- Examples of spike selection: spikes (in green) close to prototype's spikes are selected





- The prototype with maximal proximity to selected spikes determines the category (interpreting the SVM as a nearest-neighboor mechanism)

From categorization to segmentation

- Performances are preserved:

	Number of samples (Perror < 0.1)
pre-segmented sequences	**17**
auto-segmented sequences	**19**
non-segmented sequences	**25**

- Remark: this mechanism could be closely related to neural-network registration
  - Spikes with no correspondent in the prototype are ignored.
  - Proximity is related to the *co-incidence* of two spikes as developed in e.g. SpikeNet Thorpe et al [2001].

Could early visual processes be sufficient to label motion?

- Yes, … but
  - A realistic temporal spiking retina model has to be used: Wohrer etal [2005]
  - Feature vector representation must correspond not a ``vector ´´ but to neuronal map outputs
  - Validation on other databases has to be performed

- The role of feedbacks from early-vision categorization (to e.g. segmentation mechanisms) has to be further explored to better understand the role of such powerful mechanisms.

Goals, w.r.t. the State of the Art

• Propose an automatic system which will be sensitive to small human action differences

→ Tests on walking and marching sequences

• Start from raw videos and propose an approach to extract features, based on information coding mecanisms.

→ "early visual processes"

Keyword is rank order coding

• Use pieces of local trajectories (i.e. motion) to discriminate actions.

Spike based coding schemes

- From the stimulus input a *very high dimensional array* of ``internal values'' is computed.
- Only the *most relevant* values are taken into account, data classification, being based on a sparse representation.
- Distinguishable values are defined in a *bounded range with a minimal step*, thus separated by a margin limiting the bias.

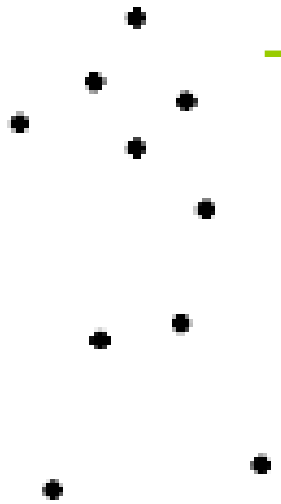
In biological systems these properties emerge from event-based coding schemes, as observed in spiking neural networks.

Simulation of retinal coding

- Matching pursuit-like algorithm (Mallat [1993])

Illustration on a grey-scale image
(see e.g. Perrinet, Samuelides and Thorpe [2004])

- Extension to video coding
  - Several possibilities: The simplest one in this work
    (see also e.g. Olsehausen [2003])



Thank you for your attention

To know more:

www-sop.inria.fr/odyssee