Master in Computer Vision Barcelona

Module 6: Video Analysis

Lecture 1.a: Introduction

Lecturer: Montse Pardàs

Instructors

Javier Ruiz Hidalgo
j.ruiz@upc.edu
(coordinator)

Montse Pardàs
montse.pardas@upc.edu

Ramon Morros
ramon.morros@upc.edu

Federico Sukno
federico.sukno@upf.edu

Gloria Haro
gloria.haro@upf.edu

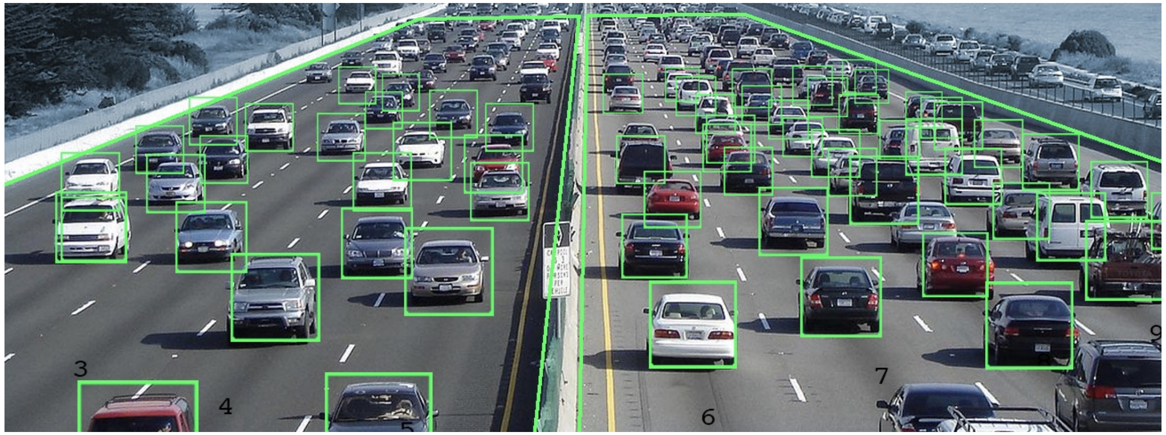
Contents – Theory classes

- **Introduction (1h)** (M. Pardàs)
- **Video segmentation (3h):** Shot segmentation, change detection, background subtraction, region segmentation (M. Pardàs)
- **Motion estimation (3h)** (R. Morros)
- **Tracking (3h):** Bayesian tracking (Kalman filter, Particles filters) (R. Morros)
- **Recurrent Neural Models- Attention Models (2h)** (F. Sukno)
- **Transformers (2h)** (R. Morros)
- **Neural architectures for video (1h)** (R. Morros)
- **Video Object Tracking & Segmentation (1h)** (R. Morros)
- **Self-Supervised learning. Multimodal Architectures(2h)** (G. Haro)
- **Recognition (2h):** Activity pose and gestures (J. Ruiz)

Contents – Project

Supervisors: Javier Ruiz, Ramon Morros

- Goal: To learn the basic concepts and techniques related to video sequences mainly for surveillance applications
- Applications: Road Traffic Monitoring and Advanced Driver Assistance



ROAD TRAFFIC MONITORING

Schedule

https://mcv.uab.cat/m6-video-analysis/

Evaluation

- **Final Mark** = 0.4 x **Exam** + 0.55 x **Project**+ 0.05 x **Attendance**

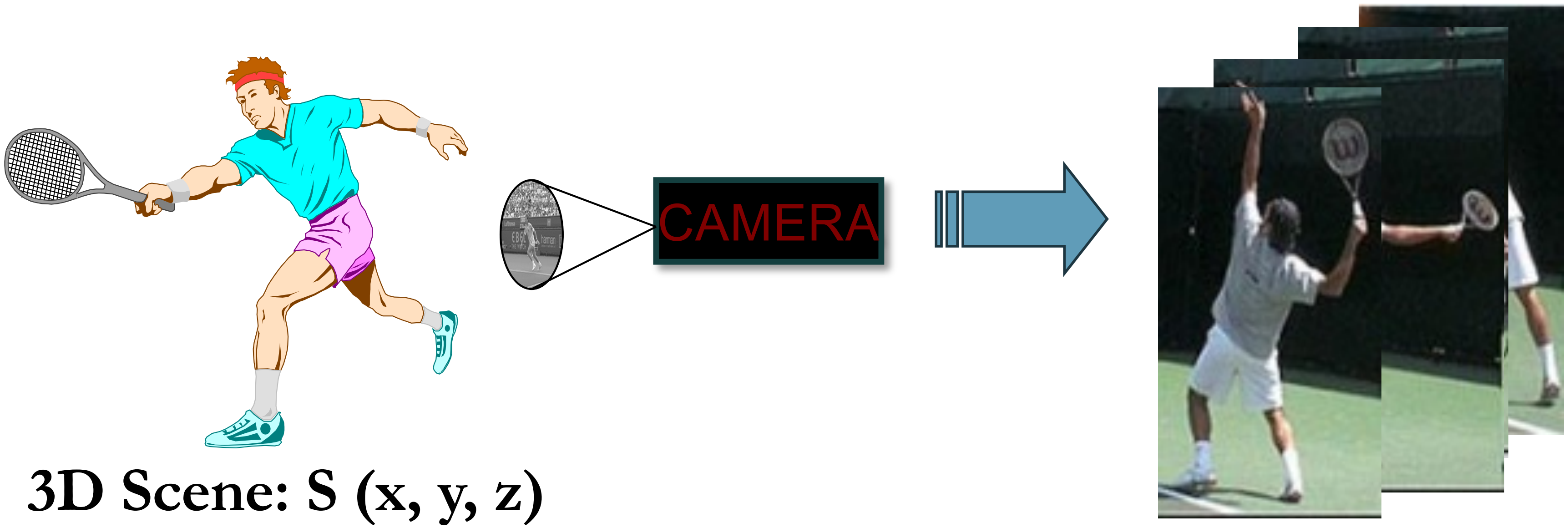
Introduction

- Perception of moving images
  - Temporal sampling frequency
  - Shaking and flickering
- 2D Motion analysis
  - Real motion, Apparent motion and Optical flow.
- Applications
- Conclusions

Perception of moving images

Temporal sampling frequency

- When a video is recorded, a 3D scene with continuous temporal evolution is **projected into a 2D space with discrete temporal sampling**.
- The frequency at which temporal samples (images) are taken has to be established taking into account the application, and the **human visual system for consumer videos**.



3D Scene: S (x, y, z)

3D Moving Scene: S (x, y, z; t)

2D Image: I (m, n)

2D Sequence: I (m, n; k)

Perception of moving images
Temporal sampling frequency

- Computer vision applications can have a wide range of sampling frequencies

Application for high frame rate (2.500 frames /second)

http://www.foxtenn.com/storage/app/media/video/Foxtenn_promo.mp4

Application for low frame rate

Perception of moving images

Shaking and Flickering

- **Motion continuity**: A moving object must have a natural representation: **image shaking** must be avoided. The human visual system can process up to 10-12 different images per second (**frame rate**). If the frame rate is increased, an illusion of continuity is created, producing **the impression of smooth motion**. This effect depends on **the display distance**.

12 frames/s

25 frames/s

50 frames/s

Perception of moving images

Motion continuity

A slow (fast) moving object requires 10 (14) images/second to be perceived with **smooth motion:**

- **Cinema:** **Frame rate**
  - Initial frame rate: (not standard) 16 images/second
  - Currently: 24 images/second
- **Television:** **Frame rate**
  - Europe: 25 images/second
  - USA: 30 images/second
- **Cartoons:** **Frame rate**
  - Slow objects: 12 (even 6!!) images/second
  - Fast objects: 24 images/second
- **Mobile phones:** **Frame rate**
  - Optional: 24/30/60 images/second

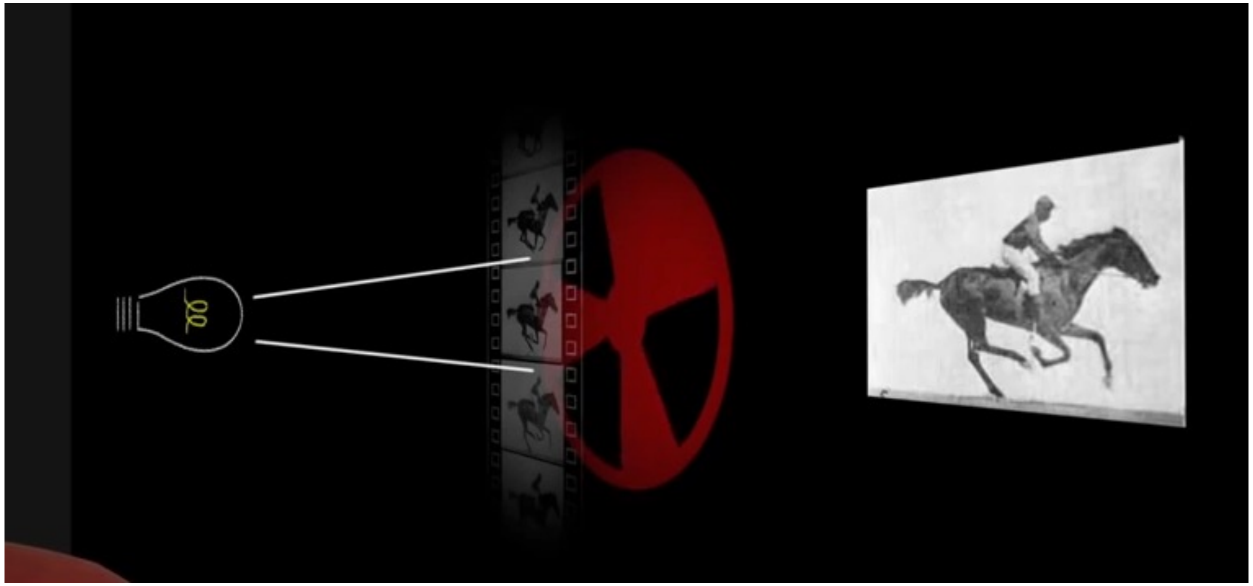
How are motion picture films displayed in TV?

Perception of moving images
Flickering

- **Brightness continuity:** An still object must be observed with constant luminosity. Low refresh rate can produce a **flickering effect** if brightness drops periodically. Depends on the environment illumination and proximity to the display.

Cinema: **Refreshment rate**

- Currently: 48 images/second **Butterfly system**



Perception of moving images
Flickering

- **Brightness continuity:** An still object must be observed with constant luminosity. Low refresh rate can produce a **flickering effect** if brightness drops periodically. Depends on the environment illumination and proximity to the display.

Cinema:	Refreshment rate	
• Currently:	48 images/second	**Butterfly system**

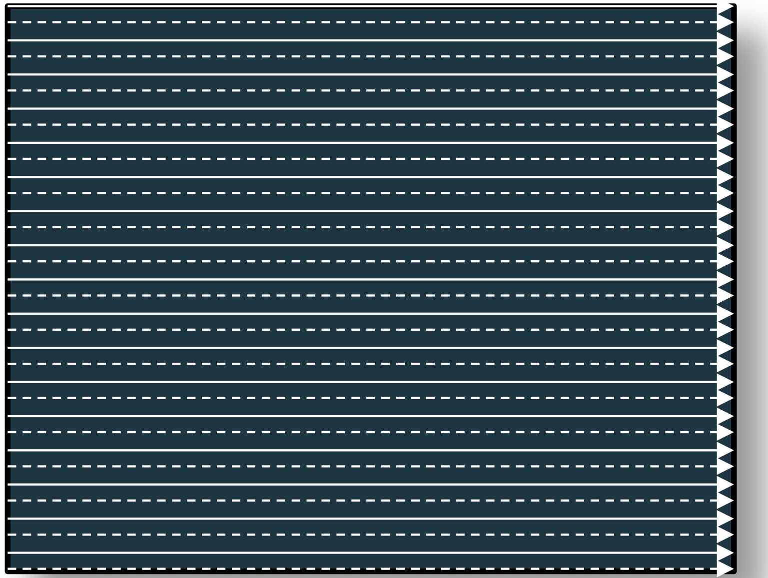
Television:	Refreshment rate	
• Europe:	50 images/second	**Interlaced**
• USA:	60 images/second	

Computers:	Refreshment rate	
• Currently:	70 … images/second Very close display	**Progressive**

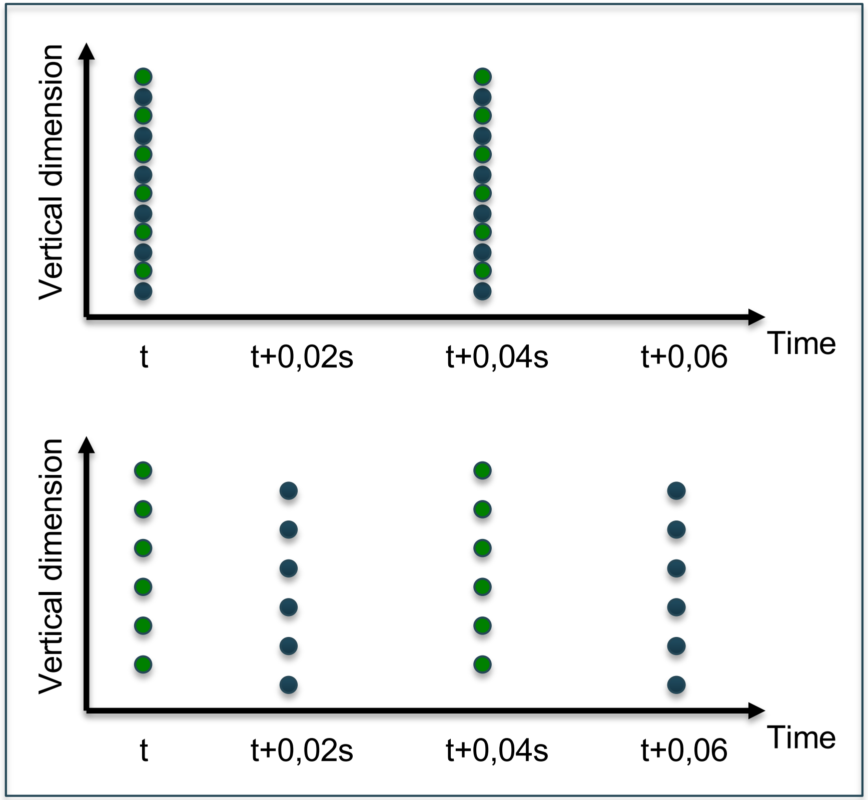
Perception of moving images

Brightness continuity: Interlaced images

- In interlaced scanning, images are **captured**, transmitted and **displayed** line by line. In order to have a **refreshment rate twice the frame rate** without **increasing the bandwidth**, odd lines and even lines are processed separately in **two different fields**.



- **Frame rate**: 25 frames/second
- Every frame is divided into 2 fields which are consecutively displayed
- **Refreshment rate**: 50 fields/second



Perception of moving images

Brightness continuity: Interlaced effect

- The fact of capturing separately odd and even lines introduces a **delay between both fields** that, if there are moving objects in the scene (or the camera is moving) creates the interlaced effect.

The interlaced effect is still important in digital television broadcasting since bandwitdth is limited and many cameras are still interlaced.

The interlaced effect is still present in digital TV.

Perception of moving images

Brightness continuity: Interlaced effect

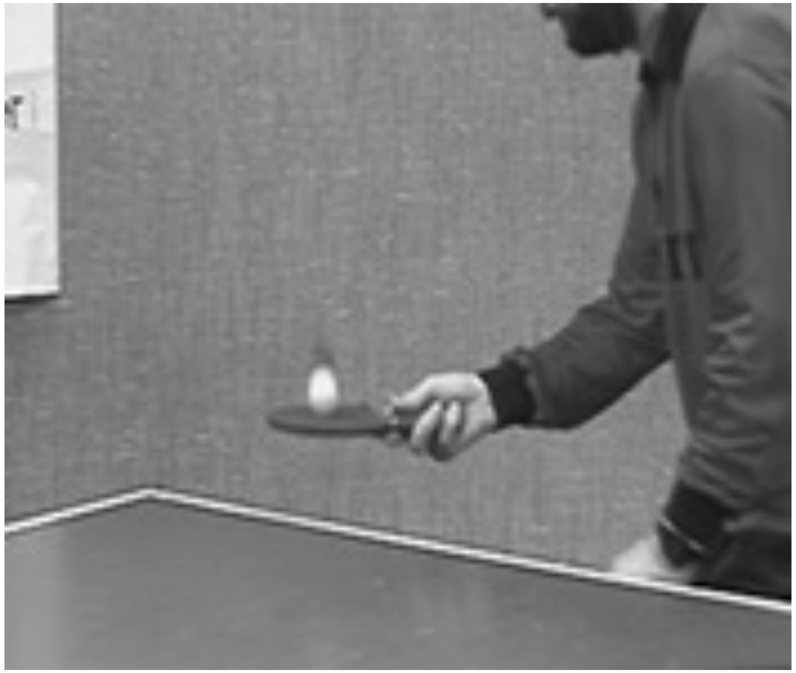
Outline

- Perception of moving images
  - Temporal sampling frequency
  - Shaking and flickering
- **2D Motion analysis**
  - **Real motion, Apparent motion and Optical flow.**
- **Applications**
- **Conclusions**

2D Motion analysis (I)

How many types of motion can you see in the scene?

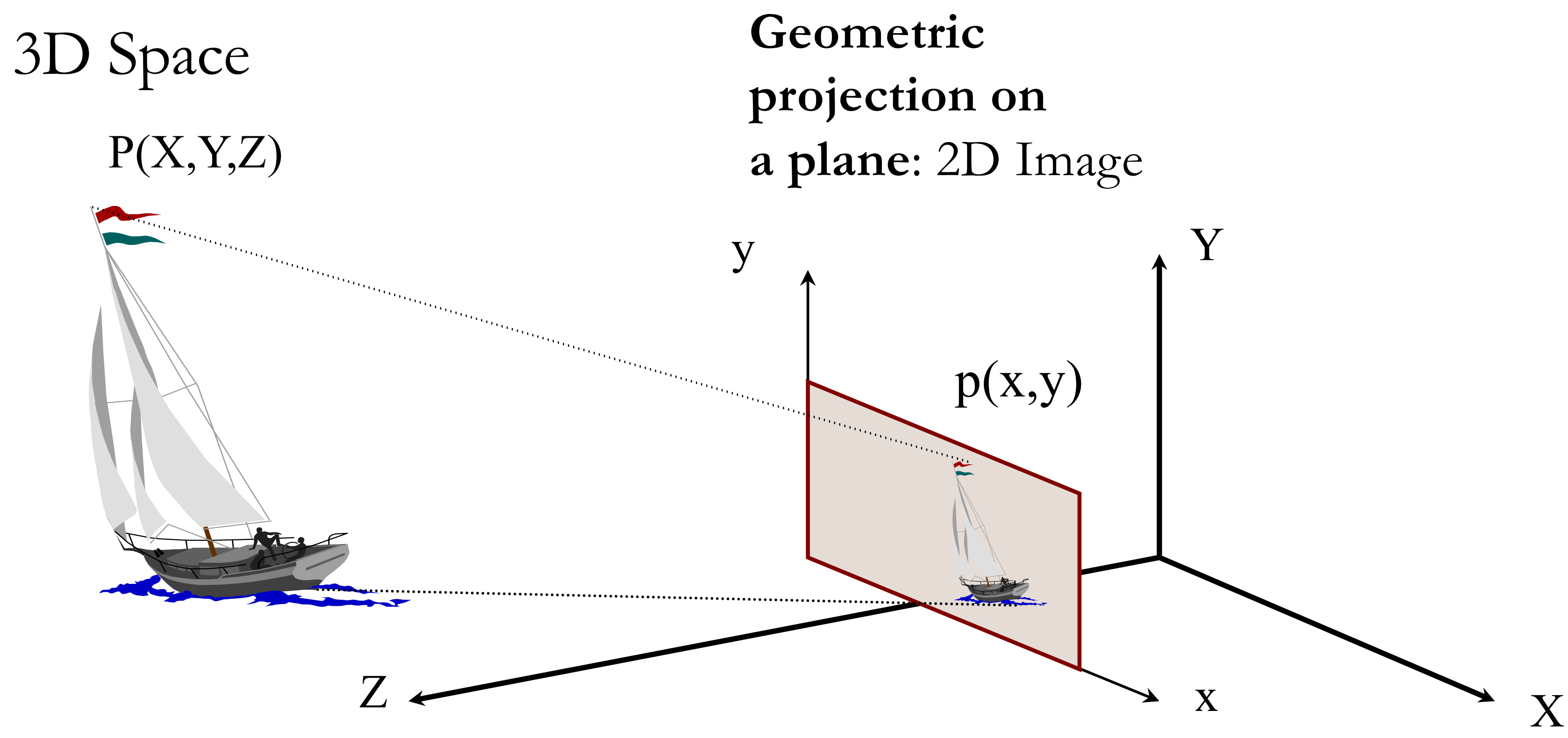
- Prompt the moving objects in the scene and **try to describe their type of motion** (Non-consecutive images are displayed to better illustrate the motion in the scene).



Frames #40 and #45 of the *Table Tennis* sequence

2D Motion analysis
Real motion

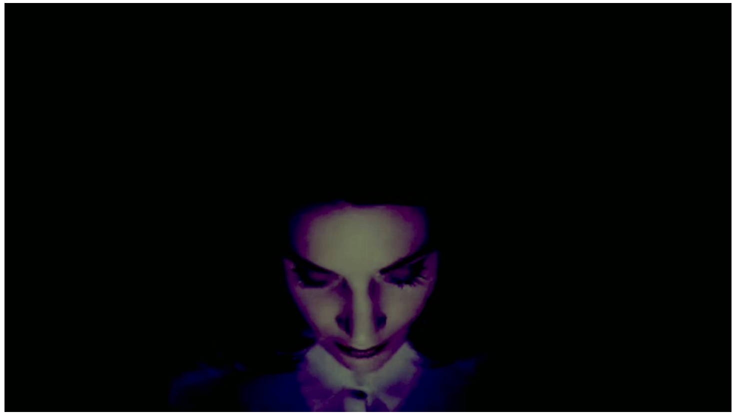
- The **relative movement** between the camera and the objects in the 3D scene leads to changes in consecutive images. Since this is the only motion that can be observed in the images, it is called **real motion**.



2D Motion analysis

Real motion, Apparent motion and Optical flow

- **Real motion** comes from the projection of the 3D moving objects of the scene onto the image plane.
  - Since the camera may be moving, we can only experiment the **relative motion** between the objects and the camera.
  - Real motion is perceived through the **pixel value changes** that appear in consecutive images.
- Pixels variations are due to
  - **real motion** as well as
  - **illumination** changes,
  - **reflections**,
  - **noise**
  - This creates **the apparent motion**.



2D Motion analysis

Motion can be caused by variations in the object, the camera or the light:

– Static camera, moving objects (surveillance)

– Moving camera, static scene (3D capture)

– Moving camera, moving scene (sports, movie)

– Static camera, moving objects, moving light (time lapse)

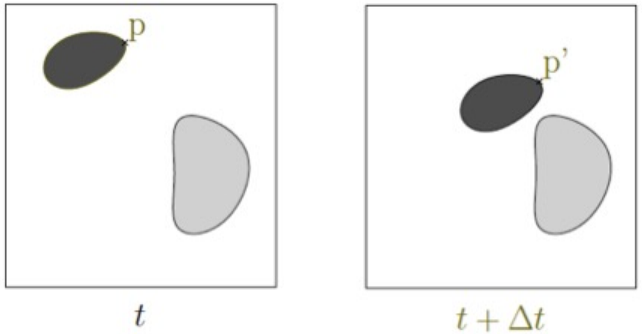
2D Motion analysis

- The mathematical study of the apparent motion allows the computation of the **optical flow**, which is a numerical modeling of the apparent motion that tries to capture the real motion.

- Solve the correspondence p-p':

$$p = (x_1, y_1, t_1); \quad p' = (x_2, y_2, t_2);$$

$$u = \frac{x_2 - x_1}{t_2 - t_1}; \quad v = \frac{y_2 - y_1}{t_2 - t_1}; \quad d_x = x_2 - x_1; \quad d_y = y_2 - y_1$$



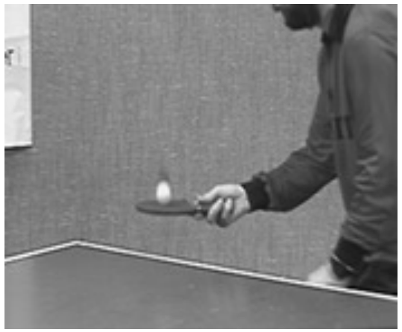
- Motion can also be recovered through **Feature-tracking**: Extract visual features (corners, textured areas) and "track" them over multiple frames

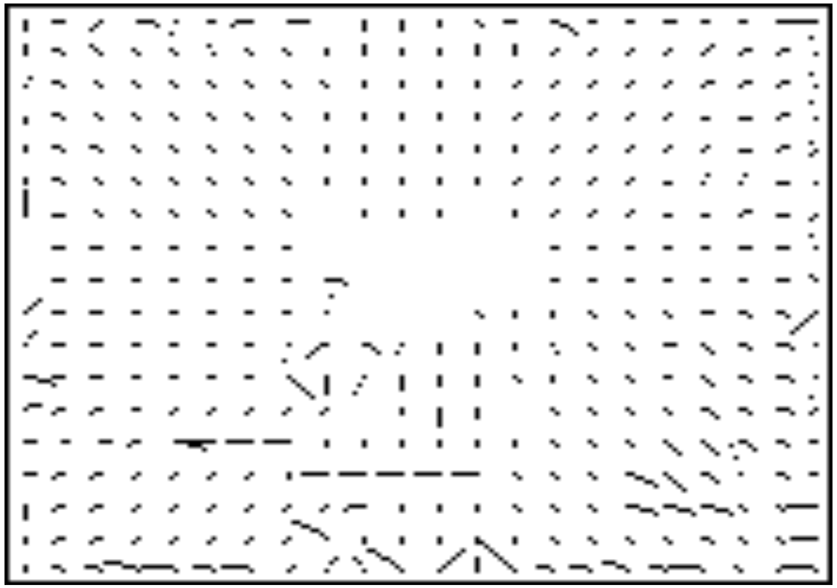
2D Motion analysis

Optical flow

- Recover image motion at each pixel from spatio-temporal image brightness variations
- The optical flow is represented by a **motion vector field**.

- For representation purposes, the optical flow shown below only presents **one motion vector for every 8x8 pixels**



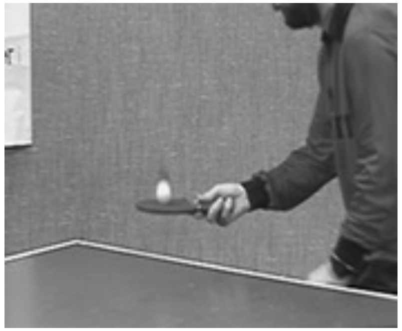
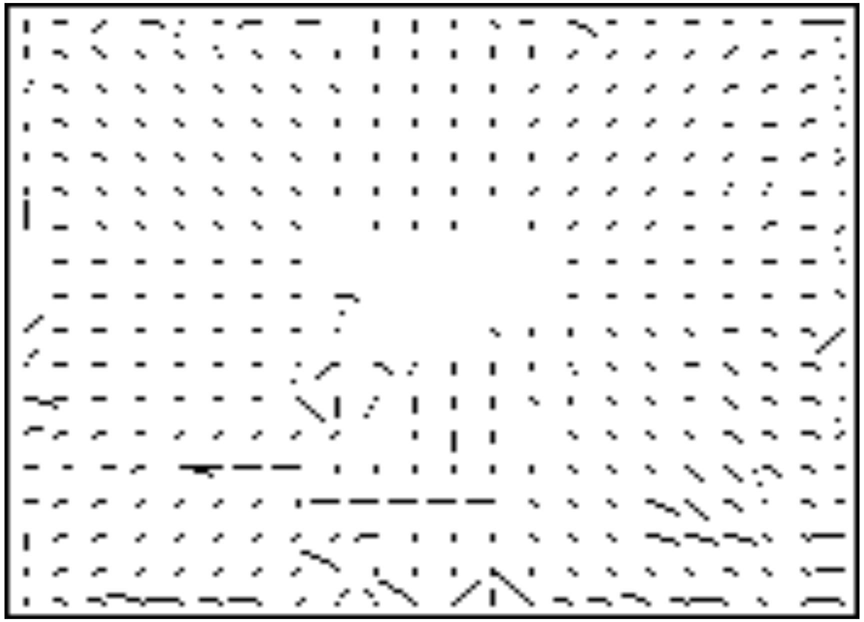


2D Motion analysis

- $\vec{D}(\vec{r})$ is the **vector function** that contains the optical flow values. It associates a **displacement vector** to each pixel
- Hypothesis for optical flow computation:

$$I(x,y,t) = I(x',y',t-\Delta t) = I(x-d_x(x,y), y-d_y(x,y), t-\Delta t)$$

$$\boxed{I(\vec{r},t) = I(\vec{r}-\vec{D}(\vec{r}), t-\Delta t)}$$

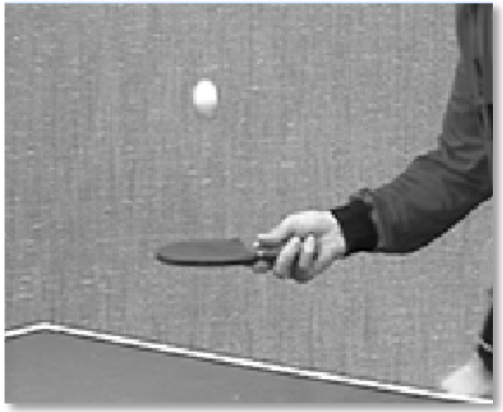




2D Motion analysis

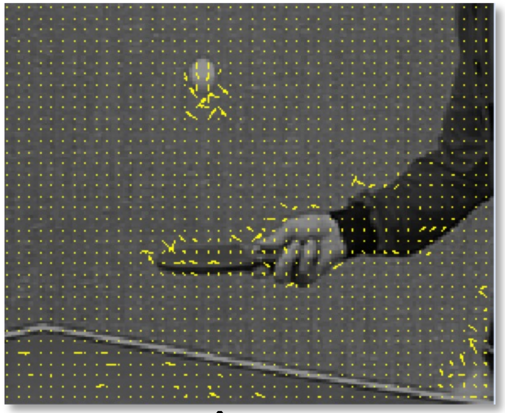
- In practice, the optical flow has to be estimated and may differ from the ideal one.

- If $\hat{\vec{D}}(\vec{r})$ denotes the estimated optical flow, the reference image to which the optical flow is applied is called the motion compensated image.

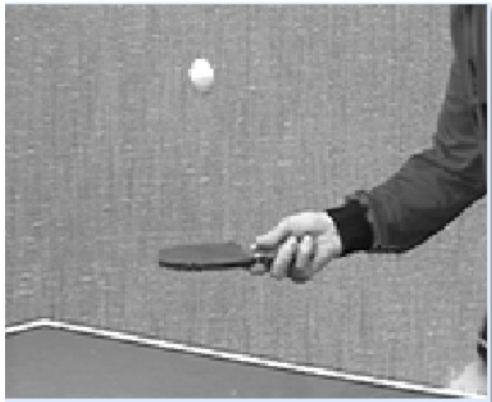
$$I^{MC}(\vec{r},t)=I(\vec{r}-\hat{\vec{D}}(\vec{r}),t-\Delta t)$$



$I(\vec{r},t-\Delta t)$



$\hat{\vec{D}}(\vec{r})$

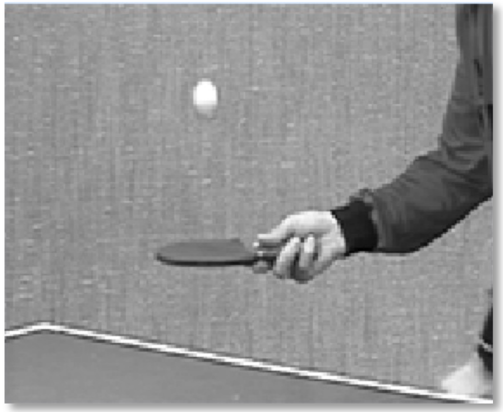


$I^{MC}(\vec{r},t)$

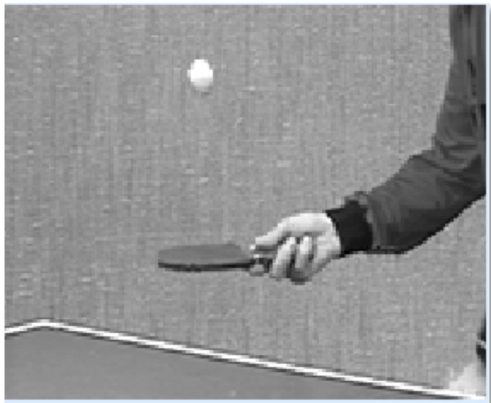
2D Motion analysis

The difference between the motion compensated frame and the real current image is called the **Displaced Frame Difference** (DFD).

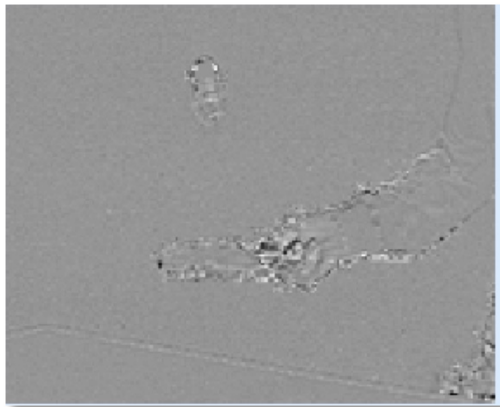
$$DFD(\vec{r},\hat{\vec{D}}(\vec{r}))=I(\vec{r},t)-I(\vec{r}-\hat{\vec{D}}(\vec{r}),t-\Delta t)$$



$I(\vec{r},t-\Delta t)$



$I^{MC}(\vec{r},t)$



$DFD(\vec{r},\hat{\vec{D}}(\vec{r}))$

Outline

- Perception of moving images
  - Temporal sampling frequency
  - Shaking and flickering
- 2D Motion analysis
  - Real motion, Apparent motion and Optical flow.
- **Applications**
- **Conclusions**

Applications of video analysis

- Biomedical Applications
- Video Coding
- Surveillance
- Human computer interaction and entertainment
- Video indexing and search
- Sports analysis
- Other applications

Applications

Biomedical applications

- Study of human male sterility caused by spermatozoa low density and/or low mobility.

Computation of the spermatozoa's mobility (Dexeus – UPC)

Detection of the individual element motion and global motion analysis

Applications
Biomedical applications

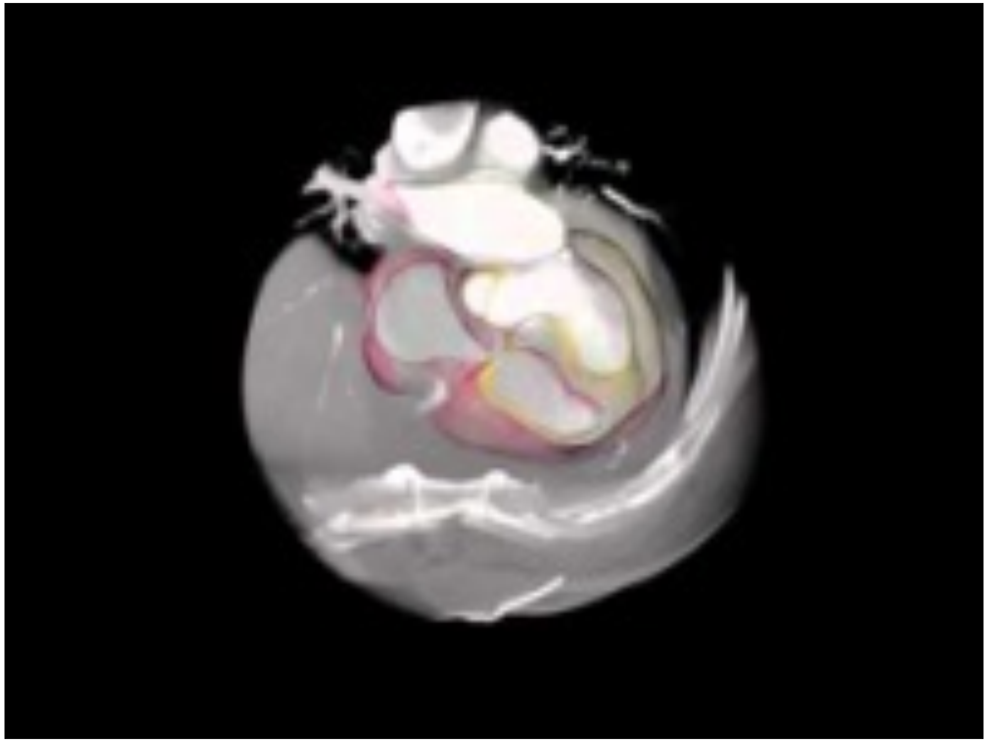
- Study of the stroke volume: Volume of blood pumped from one ventricle (left ventricle, in this case) of the heart with each beat.

Computation of the heart parameters (ESA – UPC)
Global motion of a structure

Applications
Biomedical applications

Computation of the heart parameters (ESA – UPC)
Global motion of a structure

Applications
Biomedical applications



Applications

Coding application

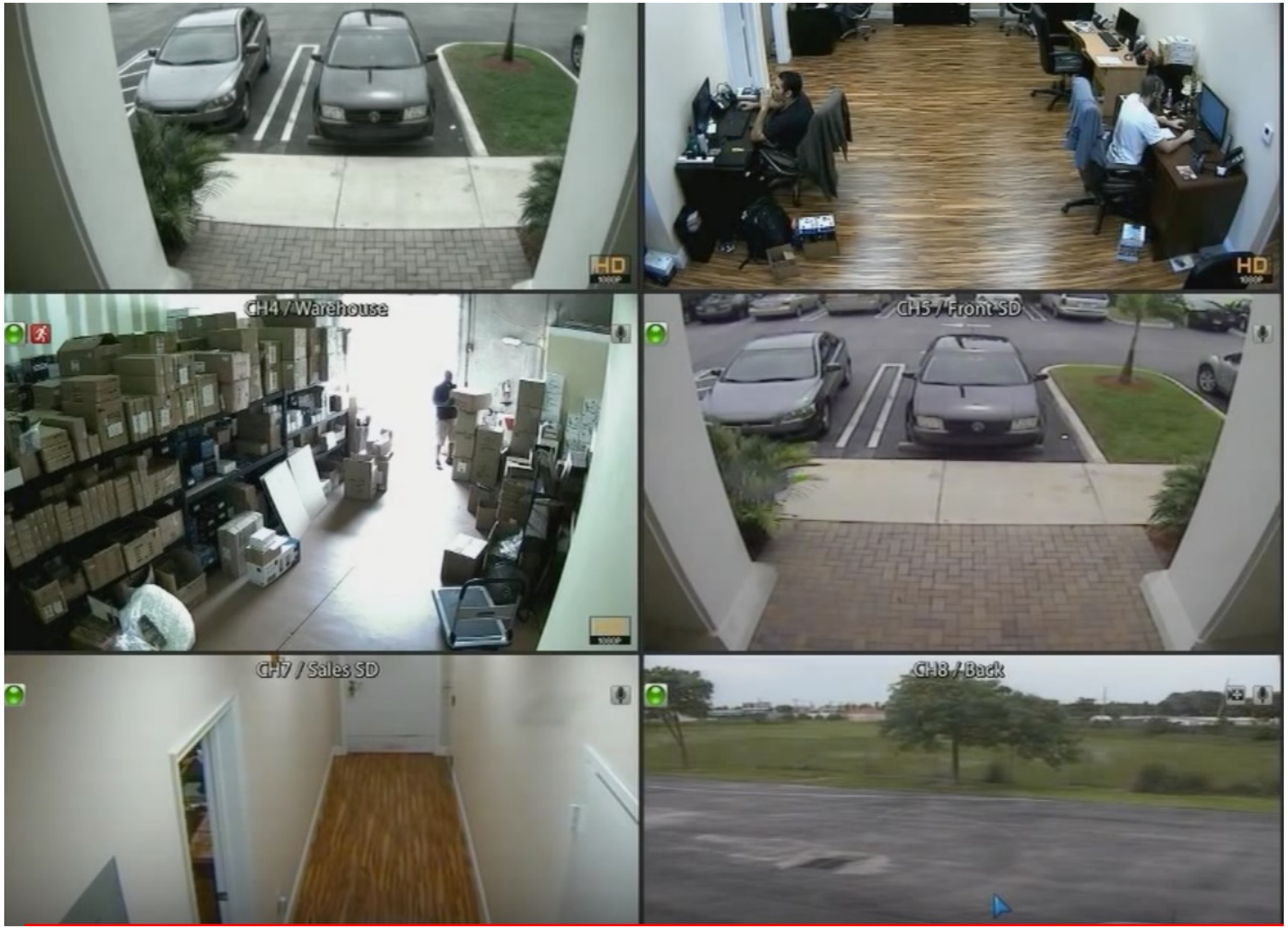
- Estimation of the motion followed by a given area.
- Motion compensation to obtain a prediction of the second image.

Exploiting the temporal redundancy of video sequences (MPEG-4 / H.264 standardization)
Motion estimation of image regions or blocks (patches) for coding applications

Applications

Security applications

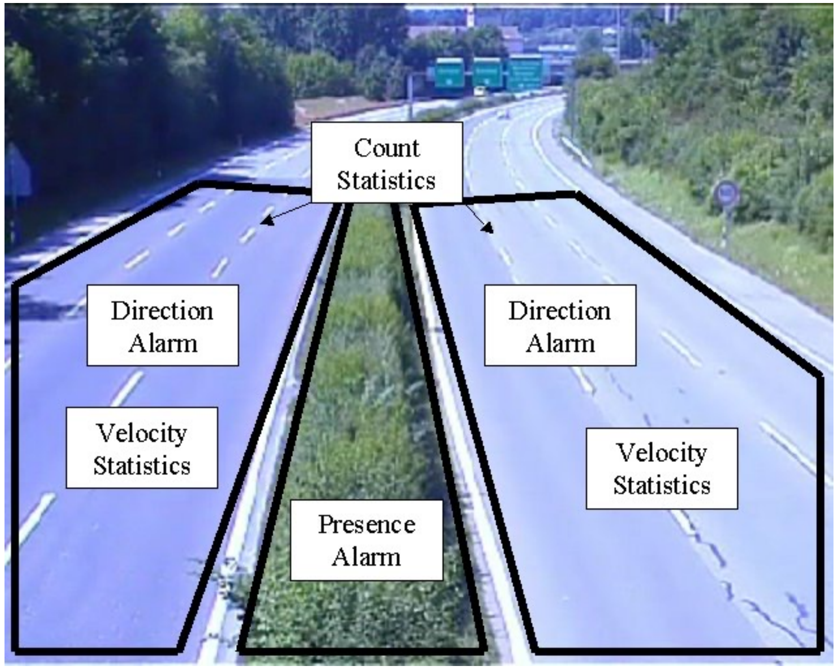
- Detection of the presence of potential intruders in areas of restricted access. Automatic monitoring of large, crowed areas such as airport installations.



Applications

Security applications

- Detection of the presence of potential dangerous behavior and objects. Analysis of the behavior of the detected people.

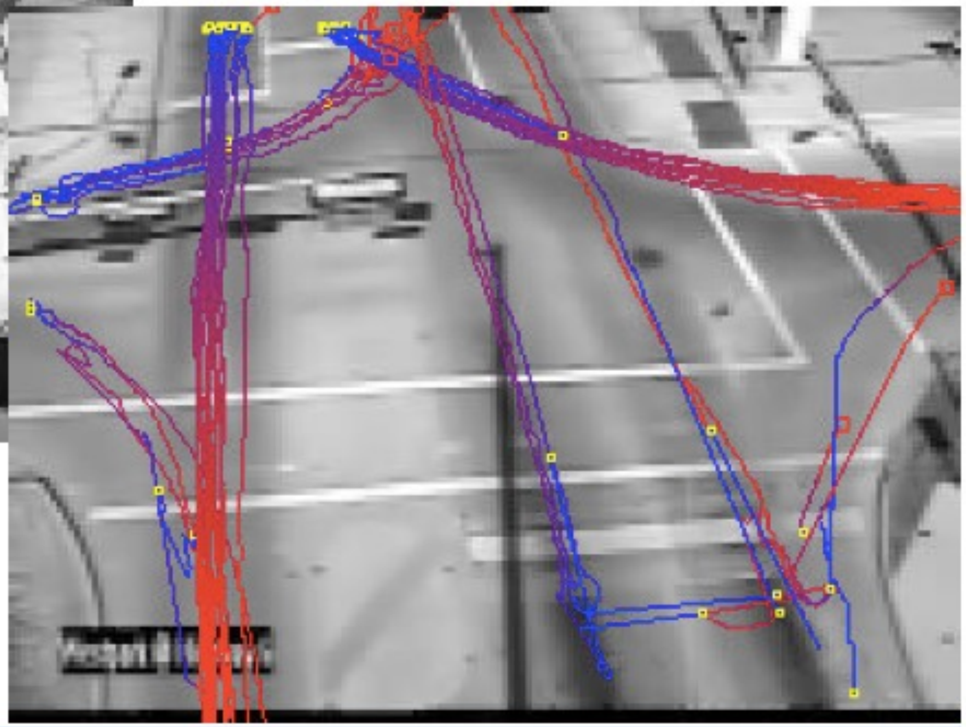


Detection of object trajectories to avoid collisions or dangerous behaviors (MODEST: UCL – EPFL)

Object tracking and motion representation in terms of trajectories

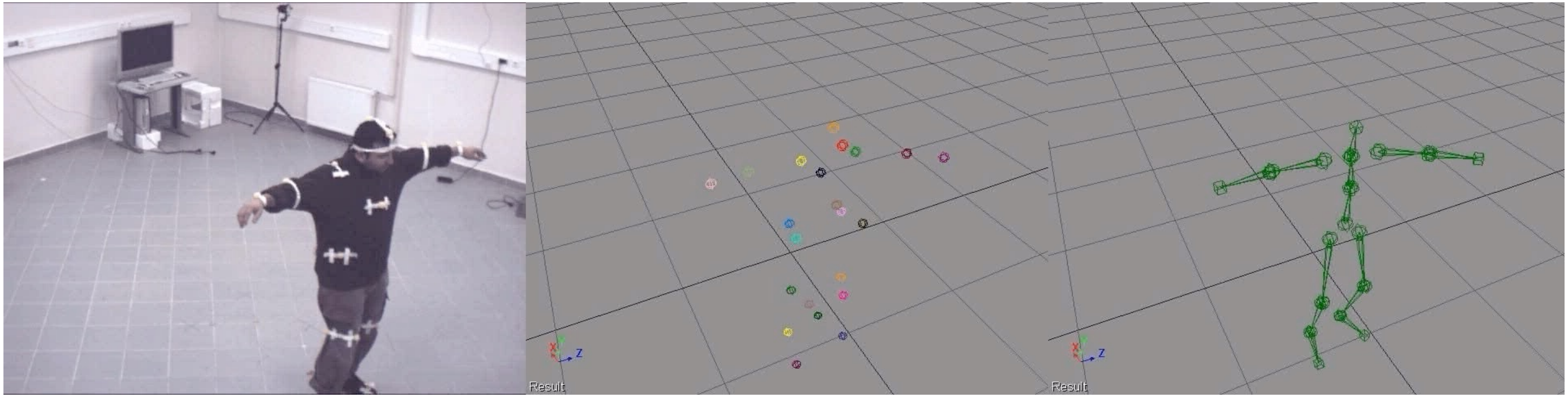
Applications

Trajectories



Applications
Modeling of specific movements

- Detection and tracking of specific markers to study their relative position and infer the global motion of the object/person.



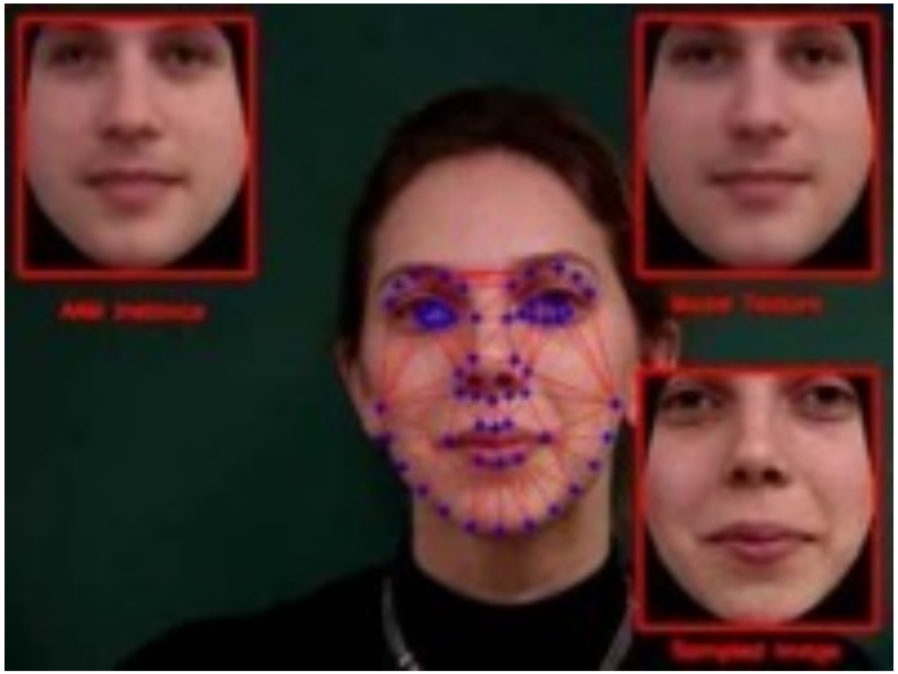
Modeling of the human motion to analyze gestural behavior (SIMILAR: Alterface - KOÇ - UPC)
Tracking of points and joint global motion analysis

Applications

Modeling of specific movements

Applications
Modeling of specific movements

- Marker-less detection and tracking of specific points to study their relative position and infer the global motion of the structure.



Modeling of the facial features motion to automatically classify facial expressions (VIDAS: LiU - UPC)

Motion analysis of specific objects and description in terms of snakes (PDEs)

Applications

Modeling of specific movements

KAIROS

KAIROS.COM

Modeling of the facial features motion to automatically classify facial expressions (VIDAS: LiU - UPC)

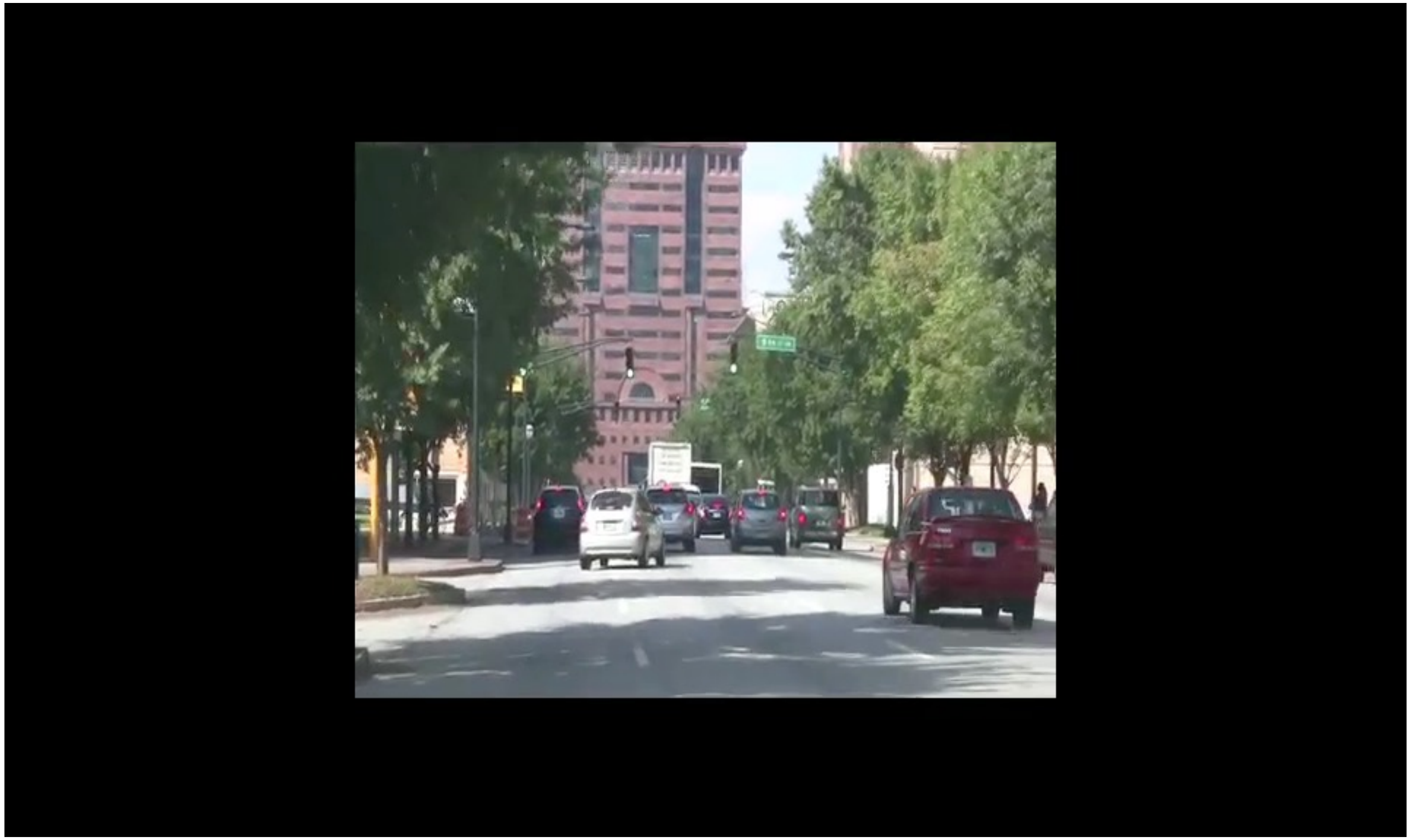
Motion analysis of specific objects and description in terms of snakes (PDEs)

Applications
Modeling of specific movements

Special effects

Applications

- Spatio-temporal region classification for indexing and search



Applications

- Semantic segmentation for autonomous driving

Applications

- Detection and tracking of lanes, traffic signs and close vehicles

Applications

- Sports analysis

Conclusions

- **Moving images perception**:
  - A minimum number of images per second is necessary to avoid **flickering** and **shaking** effects.
- **2D Motion analysis**:
  - **Real motion**: Changes due to the relative motion of the camera with respect to the objects in the scene.
  - **Apparent motion**: Actual changes observed in the image sequence (real motion, noise, illumination changes, …)
  - **Optical flow**: Displacement vectors for each pixel representing the apparent motion.

- **Applications of video analysis**