MASTER
BME

MATHÉMATIQUES ET INFORMATIQUE
Sciences
Université Paris Cité

Nicolas Loménie

Video Analysis for Bio-Medical Imaging

https://www.youtube.com/watch?v=ntk8XsxVDi0

https://www.youtube.com/watch?v=F6hdQJdWFkk

MASTER
BME

Université de Paris

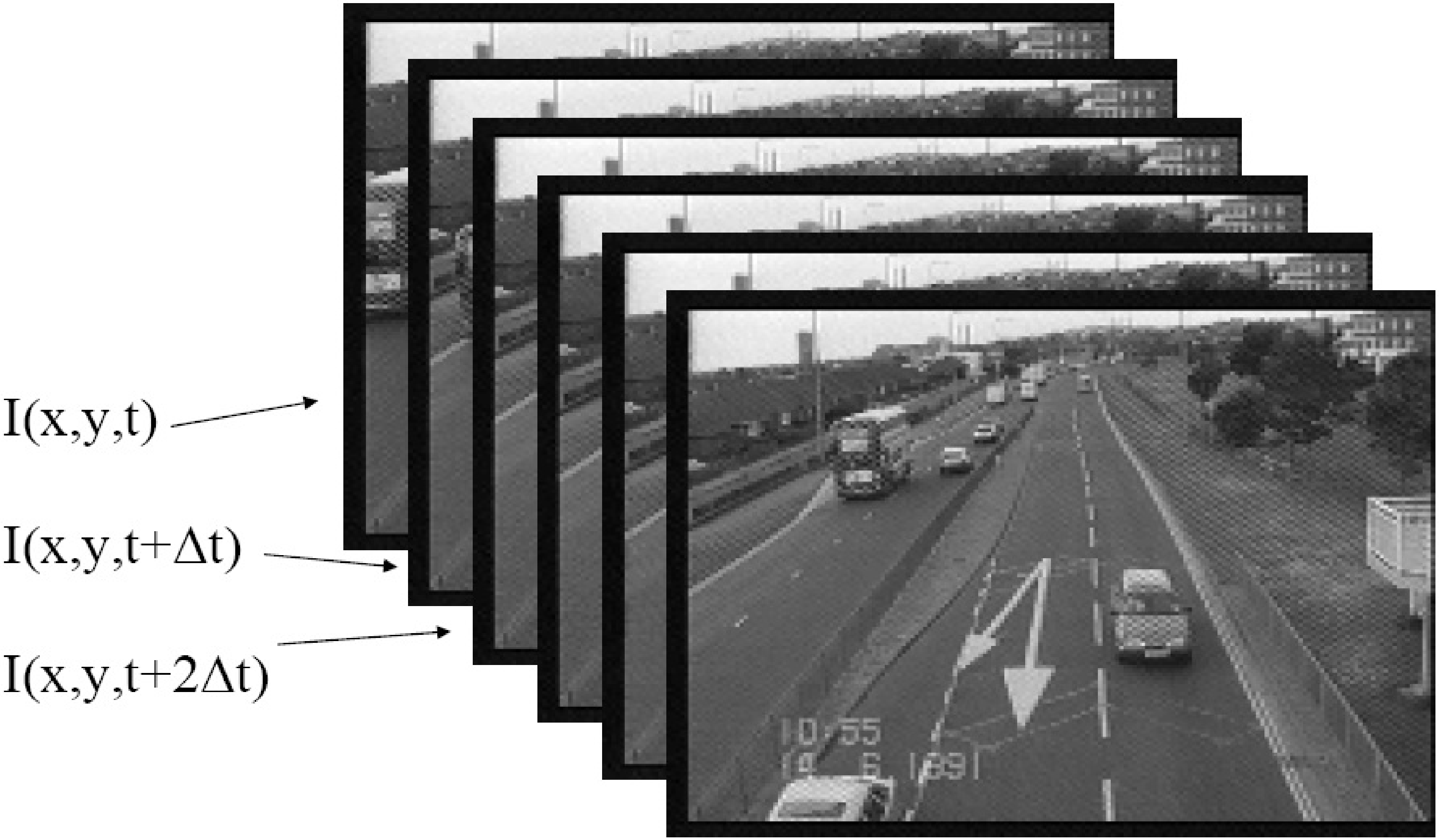
Video Analysis

Nicolas Loménie

Motion

INTRO

A global issue in Computer Vision
and Cognitive&Physiology Vision
(http://www.institut-vision.org/index.php?lang=en)



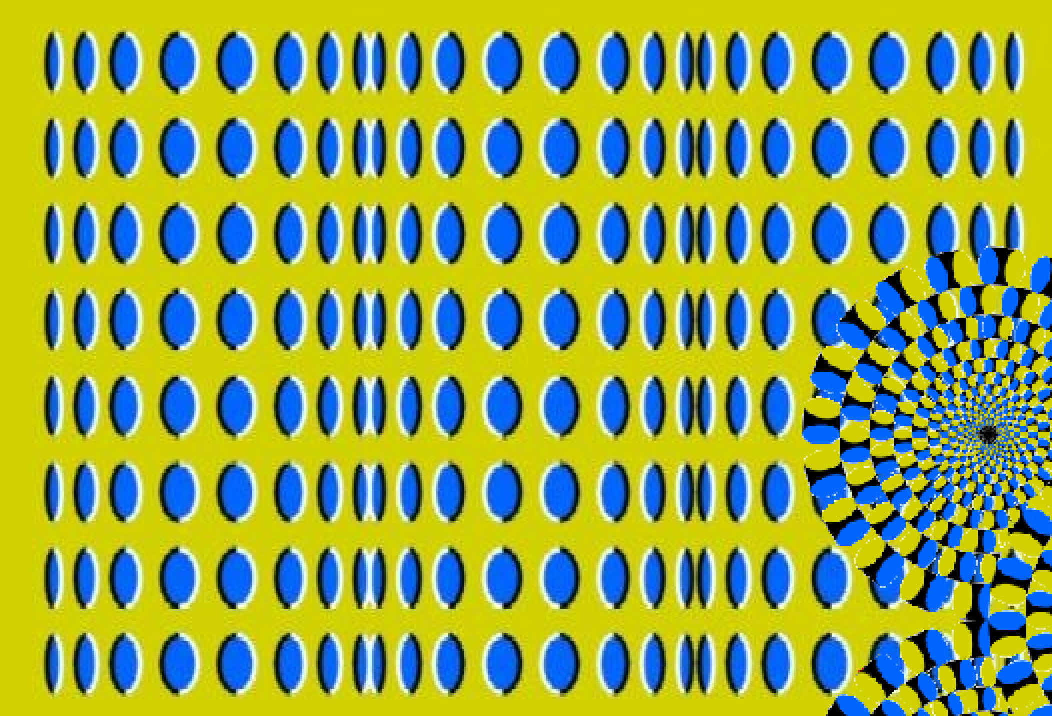
Are we working on real videos in which spatial redundancy is pre-processed (mpeg /avi files) ?

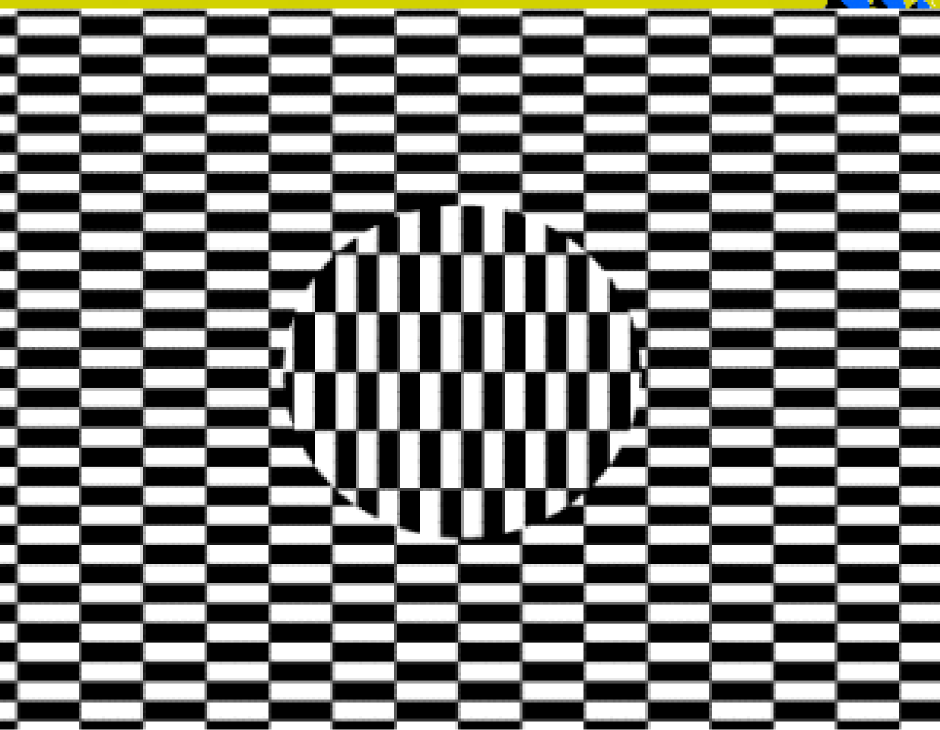
Much more on Image Sequence or stacks :

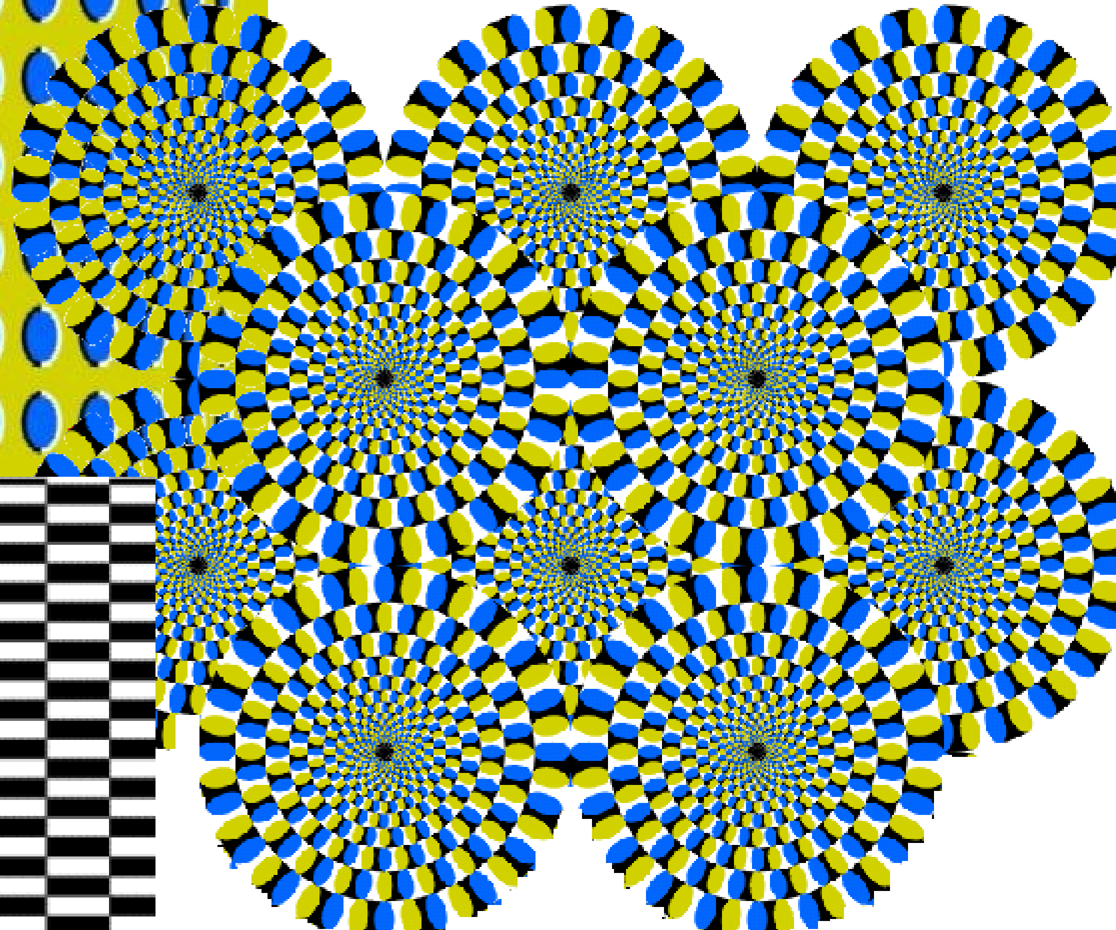
Def: An Image sequence is a series of N images, or frames, acquired at discrete intervals of time $t_k=t_0+k\Delta t$, where Δt is constant and k=0, 1, …, N-1.

Note : we need a "frame grabber" able to store frames at high rates and high volumes ("Frame rate" : Δt =1/24s or "field rate" : Δt =1/30s, time lapse Δt =1 image every hour).

FYI : retinal persistence 1/12s -> video illusion ?



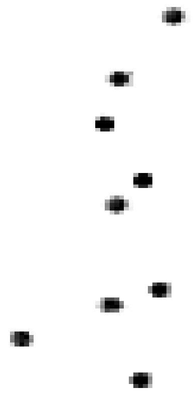




Importance of visual motion

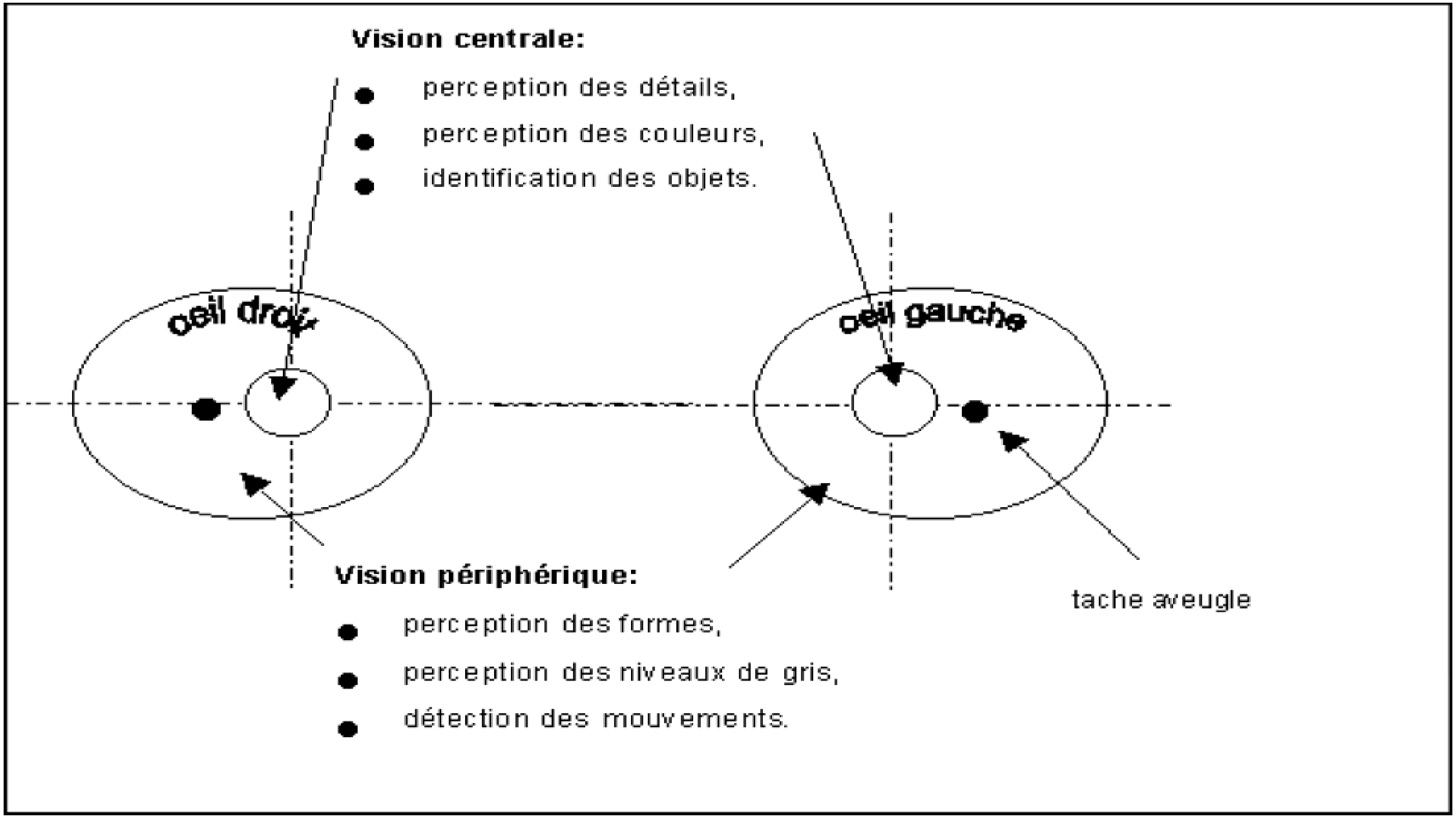
In human or computer vision, apparent motion of objects on 2D image plane is an important visual clue to understand the 3D motion AND structure in a scene

Shape from Motion topic :



Or Shape segmentation →

Visual Motion Importance

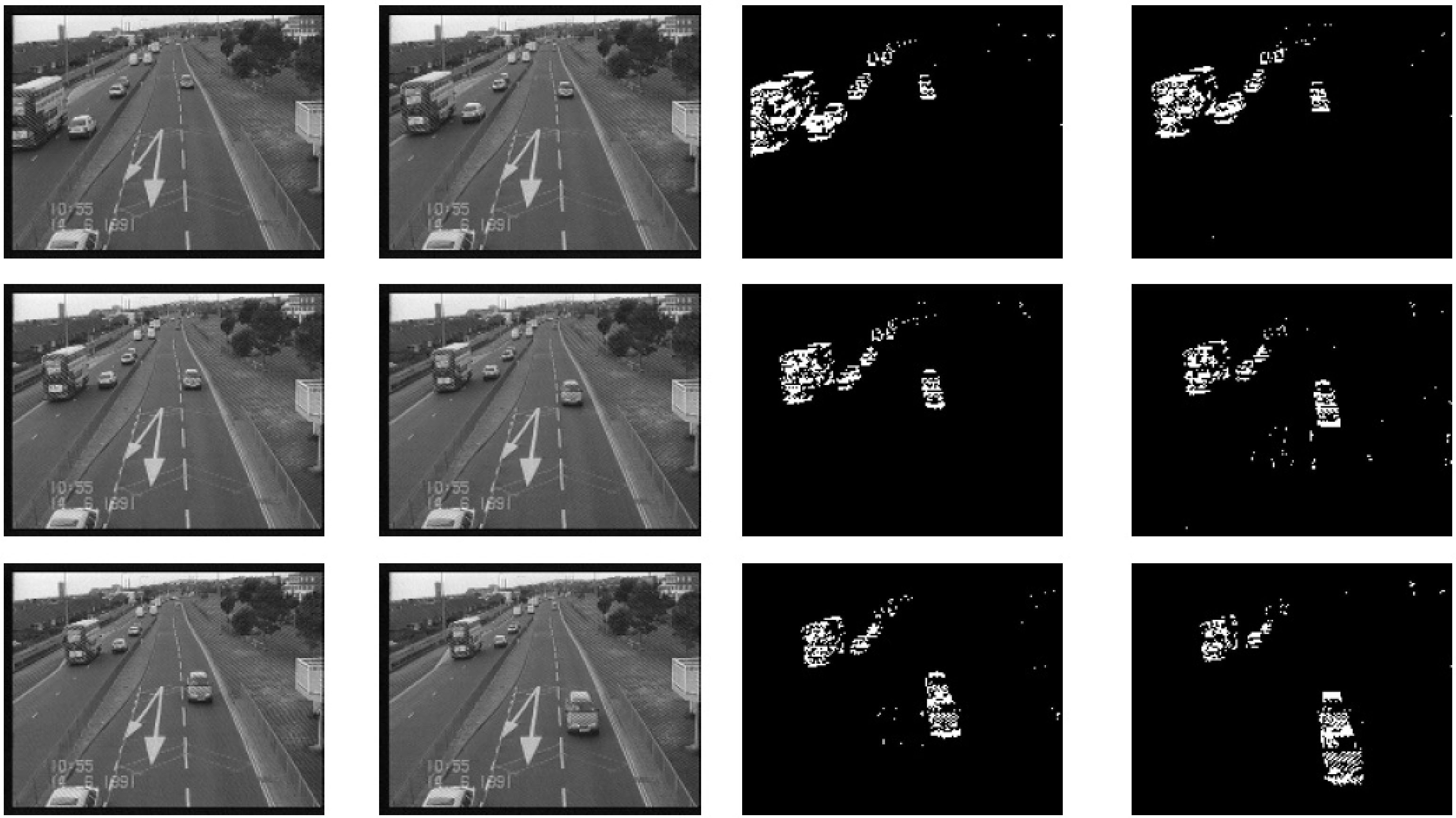


vision périphérique des vertébrés

What can we infer from 2D motion ?

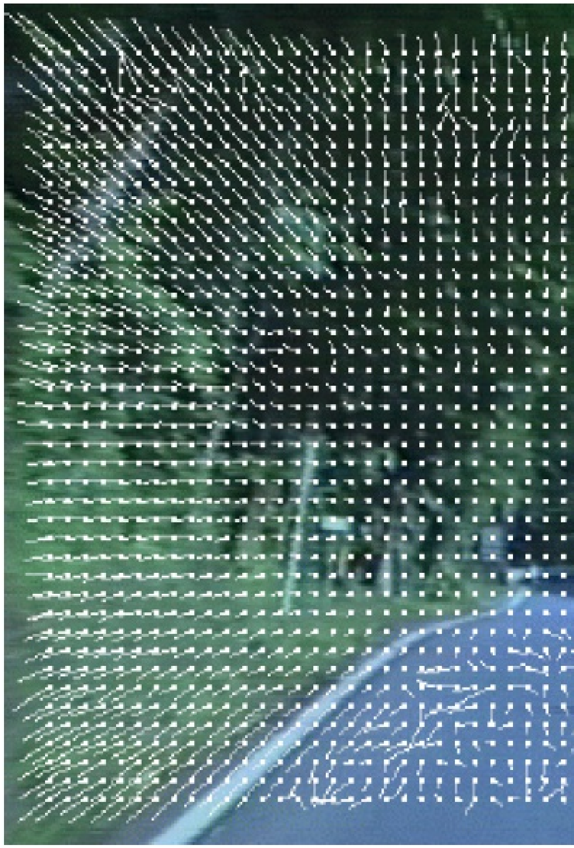
Change detection

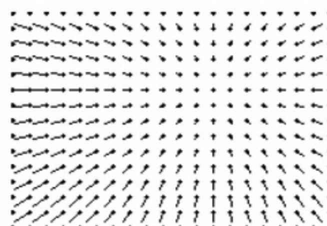
Following Traffic



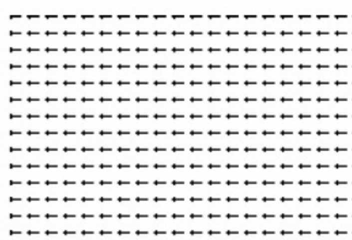
Optical-flow based analysis

Optical Flow due to camera motion

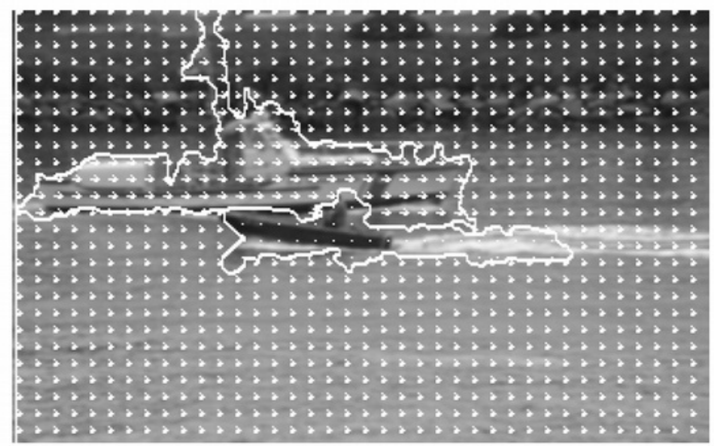




Focus of contraction



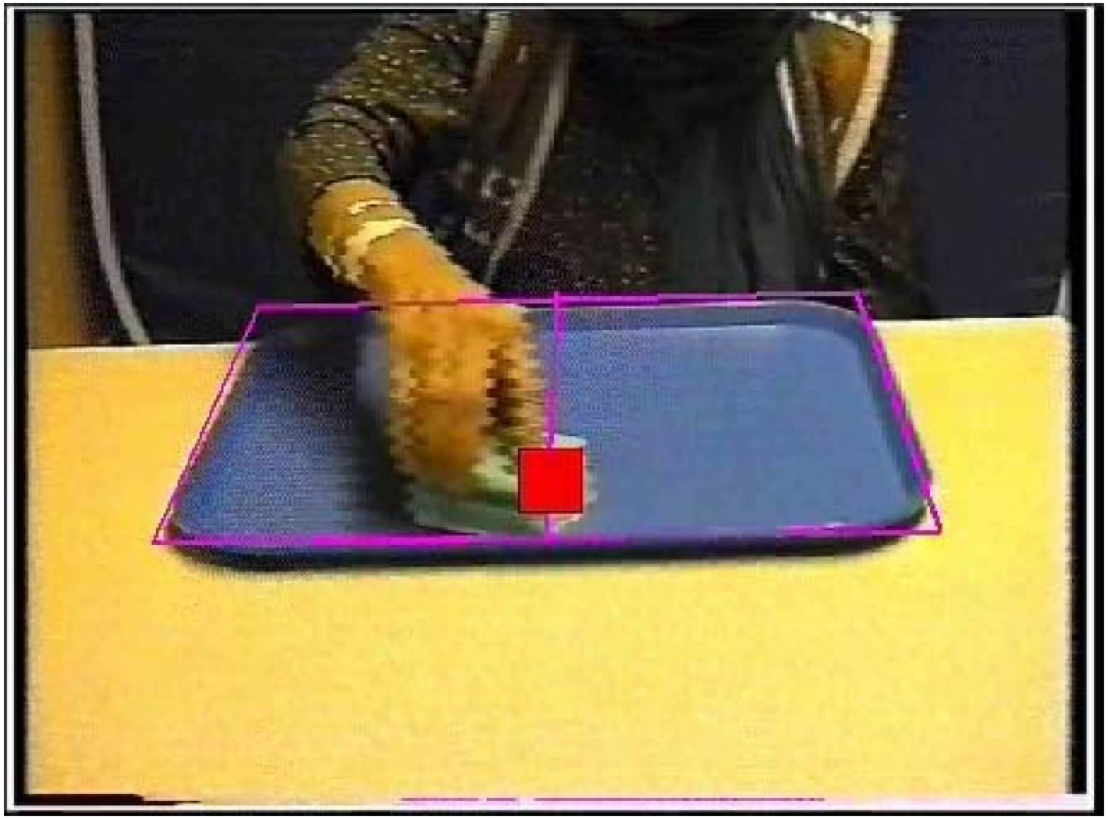
Panning



Motion-based segmentation.
Camera panning left (following smaller boat)
Larger boat moving right.

Tracking

Example: hand tracking

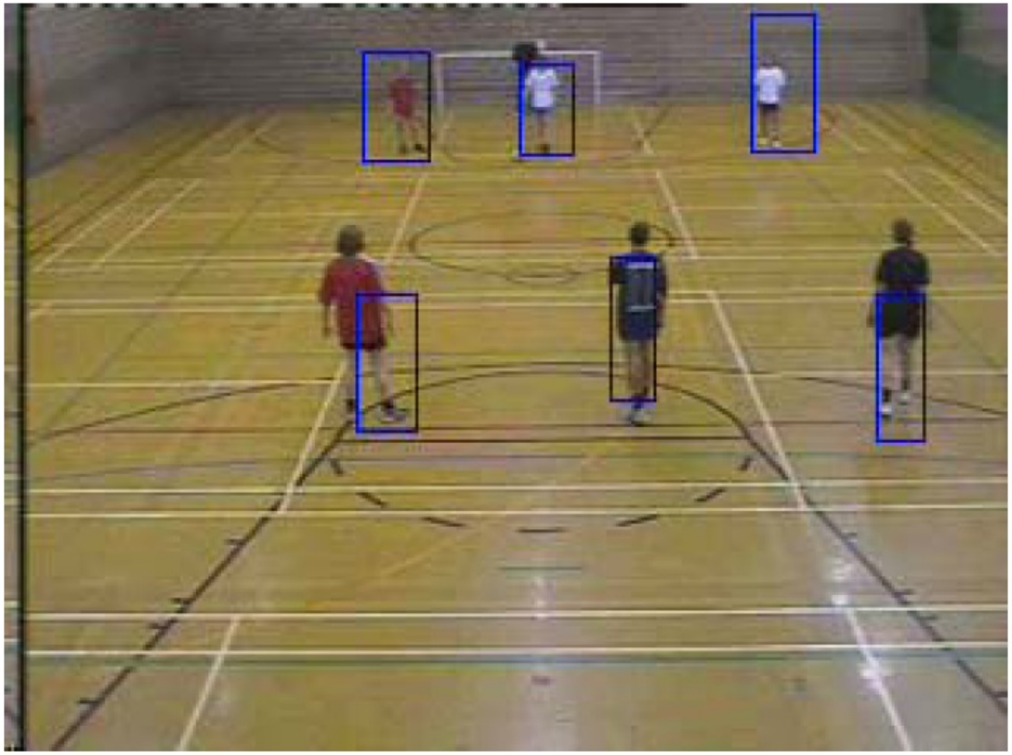


Measuring spatial neglect in stroke patients

Model-based tracking

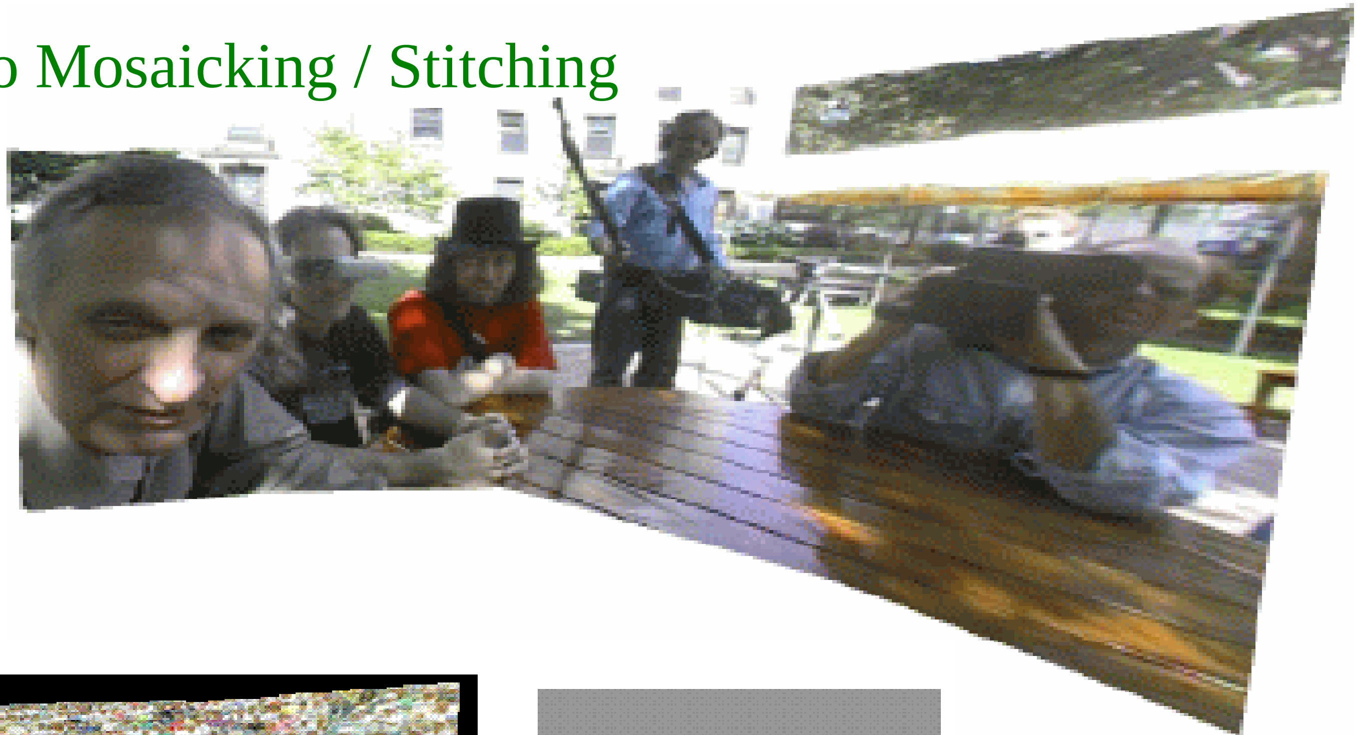
Feature-based Tracking

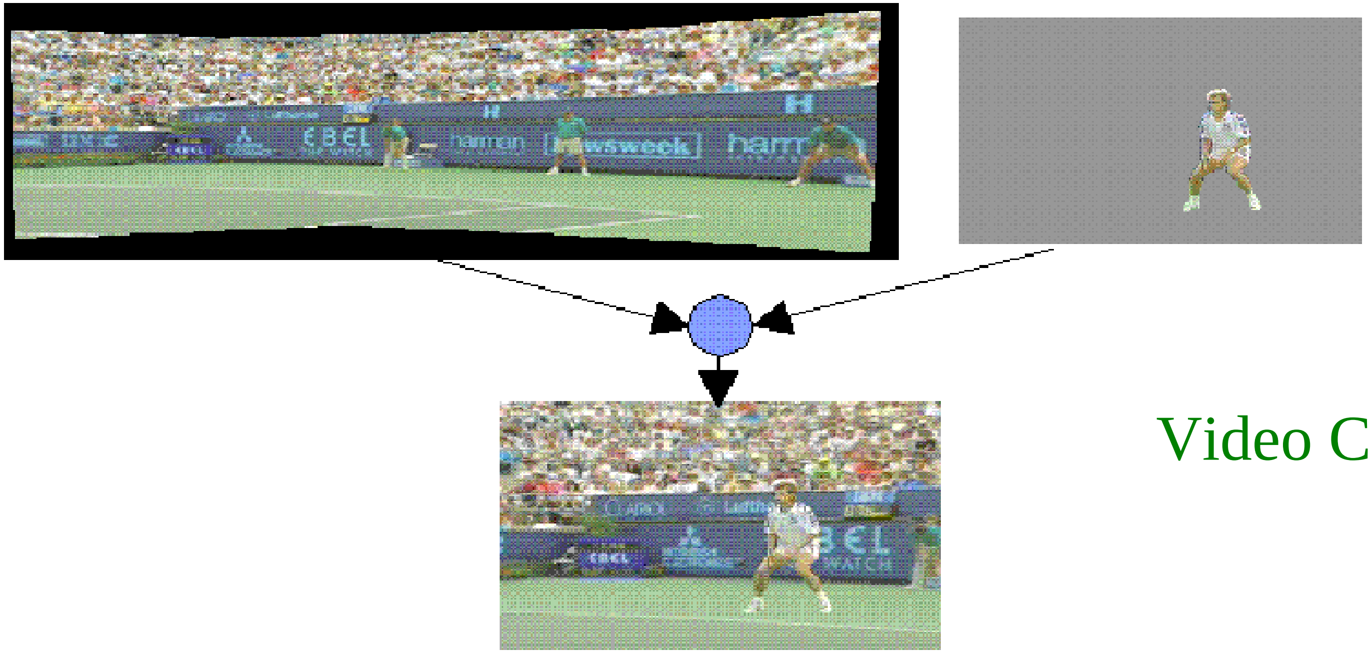
Chris Needham
and Roger Boyle,
University of
Leeds



Models based on colour
Wide variety of shapes
Multiple objects
Tracking through
occlusions
.

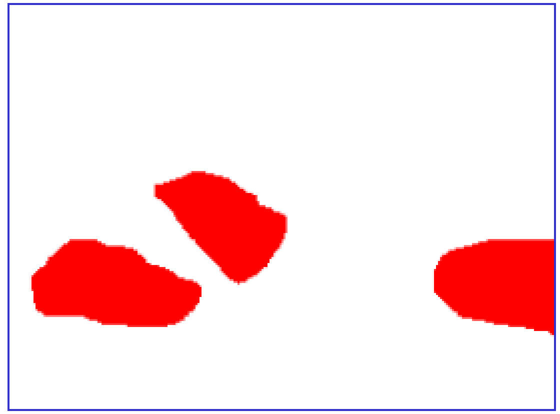
Video Mosaicking / Stitching



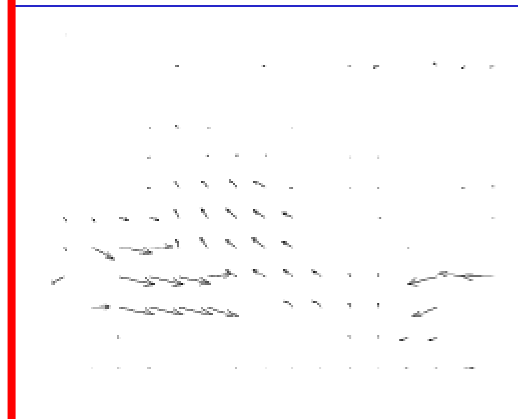


Video Compression

What kind of information to extract ?

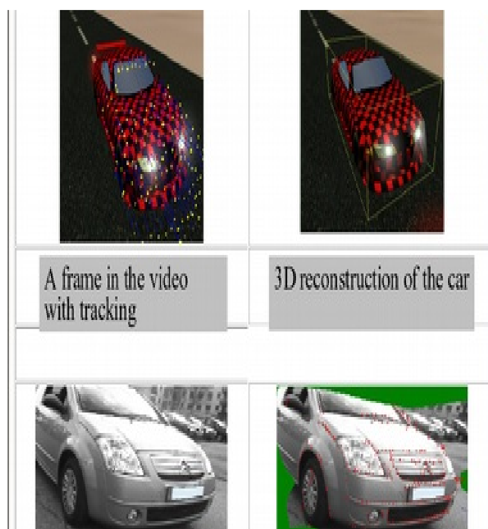


DETECTION

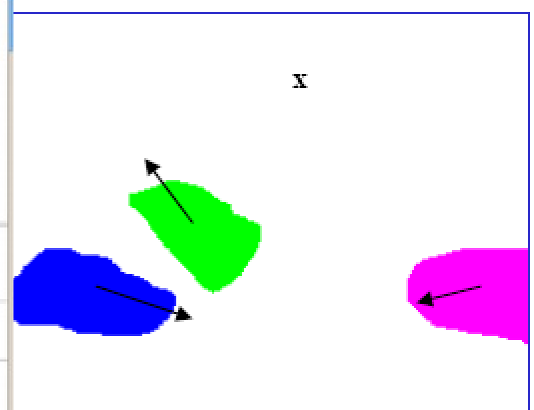


ESTIMATION

Taxi_mp4.avi



3D RECONSTRUCTION



SEGMENTATION

Conceptual level

One Core Computer Vision issue ?

The matching issue

Very general issue like image classification, indexing, etc. or Image registration (especially in the field of Medical Imaging like multi-modality).

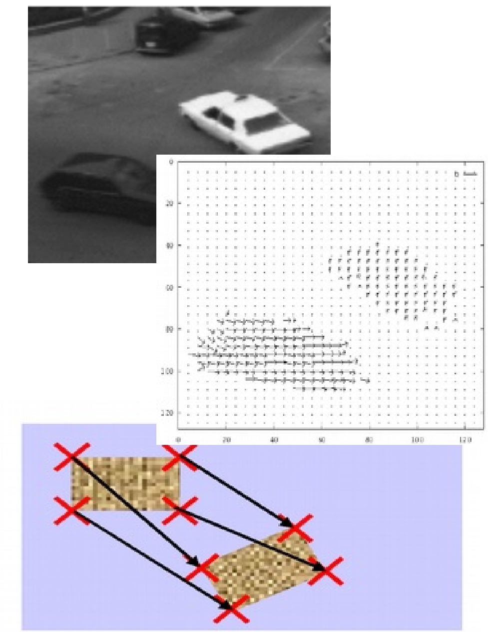
But in the case of video stream, it boils down to :

Which elements of the frame t to match with those of frame t' ?

Two theoretical frameworks:

- Differential approaches (PDE or so) : as an output, we get dense measures, i.e.. computed for every pixel of each frame. We use temporal derivatives of the signal because of the *t'-t* << *epsilon* hypothesis.

- Matching or tracking approaches (Kalman or so) : as output, we get sparse measures, i.e. computed on a subset of image feature points (SIFT, Harris etc.)



e.g. In bio-medical imaging

CREATIS Lab.-France || KULeuven-Belgium || ThoraxCenter Lab.-Netherlands

Challenge on Endocardial Three-dimensional Ultrasound Segmentation

MICCAI challenge 2014

- Overview
- Contest
- Participation
- Databases
- Evaluation
- Tutorials
- Paper Submission
- Tentative Program
- Results
- Contact

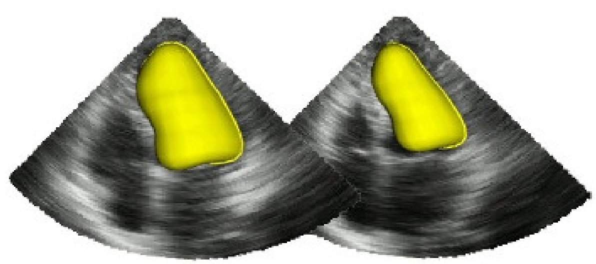
Overview

General context | **Scientific interests** | **Organizers**

The goal of this contest is to compare left ventricle segmentation methods for both End Diastolic and End Systolic phase instances. This will be done using a common database of 3D cardiac ultrasound images acquired from 45 patients and the associated manual references based on the analysis of three different experts.

Challengers will be invited to use their segmentation algorithm to automatically find LV endocardium border, in a fully automatic manner or with little user intervention related to the initialization procedure only.

All the evaluation procedure will be done fully automatically thanks to a dedicated MIDAS website.



End Diastolic Volume (EDV) | End Systolic Volume (ESV)

NEWS

The public access of the database will be made available after mid of October

22nd of July 2014
Deadline for registration

11th of July 2014
Publication-ready paper sumbission deadline

27th of June 2014
Notification of acceptance

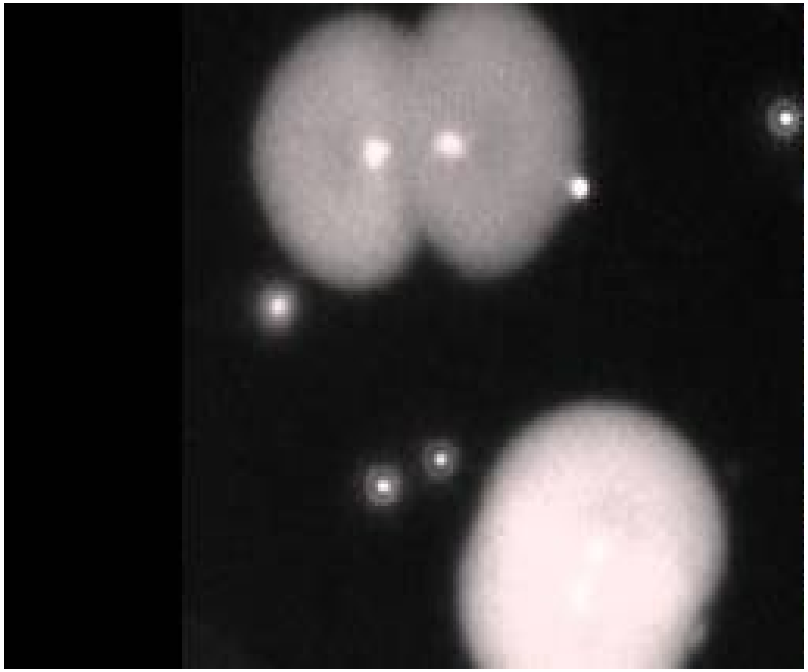
13th of June 2014
Paper submission deadline

23rd of May. 2014
1st Testing database is now available through the Midas website

http://www.creatis.insa-lyon.fr/Challenge/CETUS/index.html

C.elegans developing embryo (3D)
Waterston Lab -The George Washington University. Washington DC (USA)
Microscope: Zeiss LSM 510 Meta
Objective lens: Plan-Apochromat 63X/1.4 (oil)
Pixel size (microns): 0.09 x 0.09 x 1.0
Time step (min): 1 or 1.5

From ISBI cell-tracking challenge: http://celltrackingchallenge.net/



Sea urchin embryogenesis
Nicolas Minc Lab -Institut Jacques Monod – Université Paris Diderot (France), SPC
https://www.youtube.com/watch?v=vLXGDLDoQmc

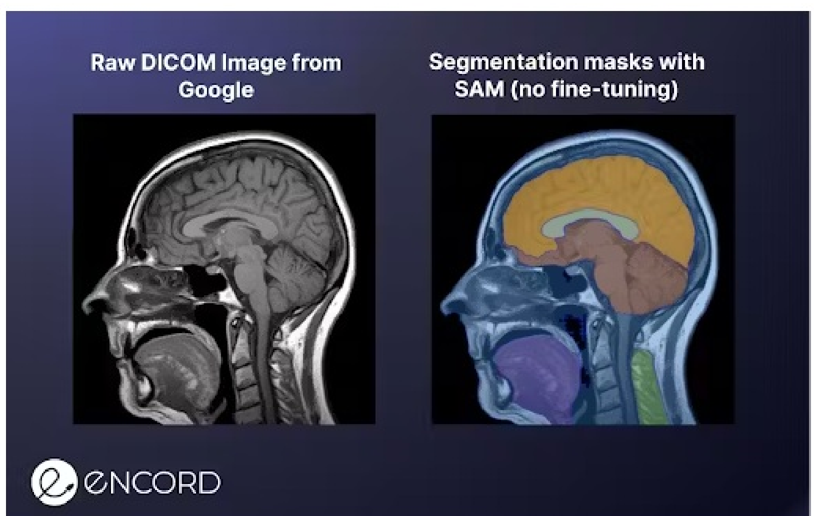
Registration is a generic issue in medical images :
- either between modalities (inter) or
- between time steps (intra).

In any case, a mixed design between matching and segmentation (the other core issue in Computer Vision but DL and GAFAM :-) https://github.com/fudan-zvg/Semantic-Segment-Anything).

https://encord.com/blog/segment-anything-model-explained/

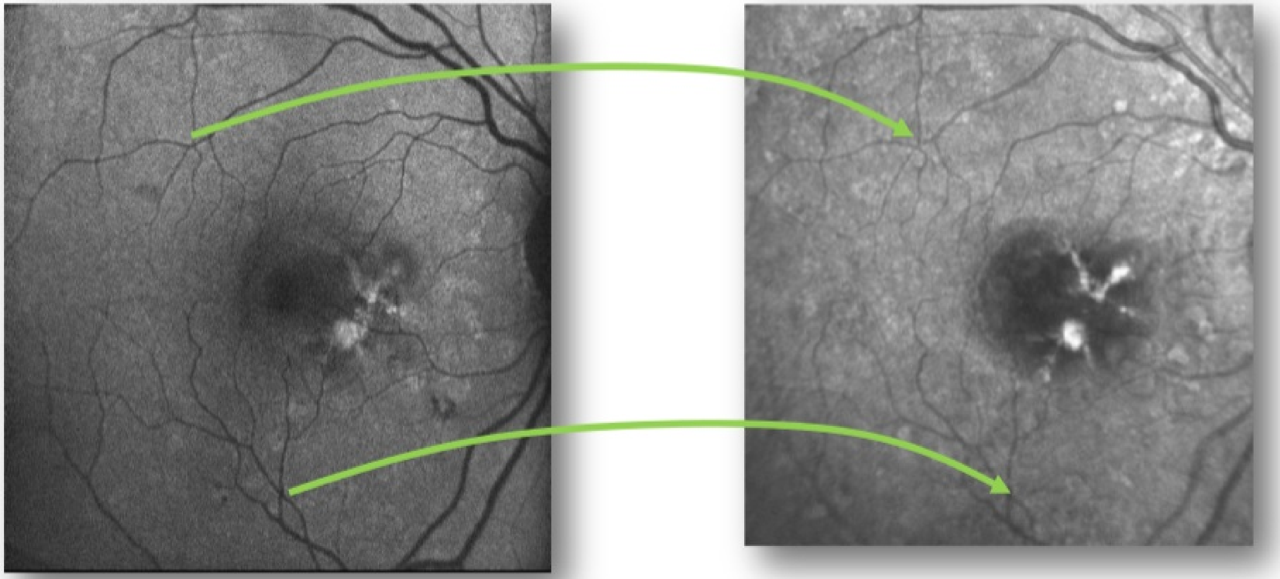
How to Use the Segment Anything Model for AI-Assisted Labeling

At Encord, we see the Segment Anything Model (SAM) as a game changer in **AI-assisted labeling**. It basically eliminates the need to go through the pain of segmenting images with polygon drawing tools and allows you to focus on the data tasks that are more important for your model.



Registration is a generic issue in medical images :
- either between modalities (inter) or
- or within the same modality between time steps (or viewpoint) (intra).

- Objective: to find a transformation or mapping V which maps the points on one image onto the points of another image

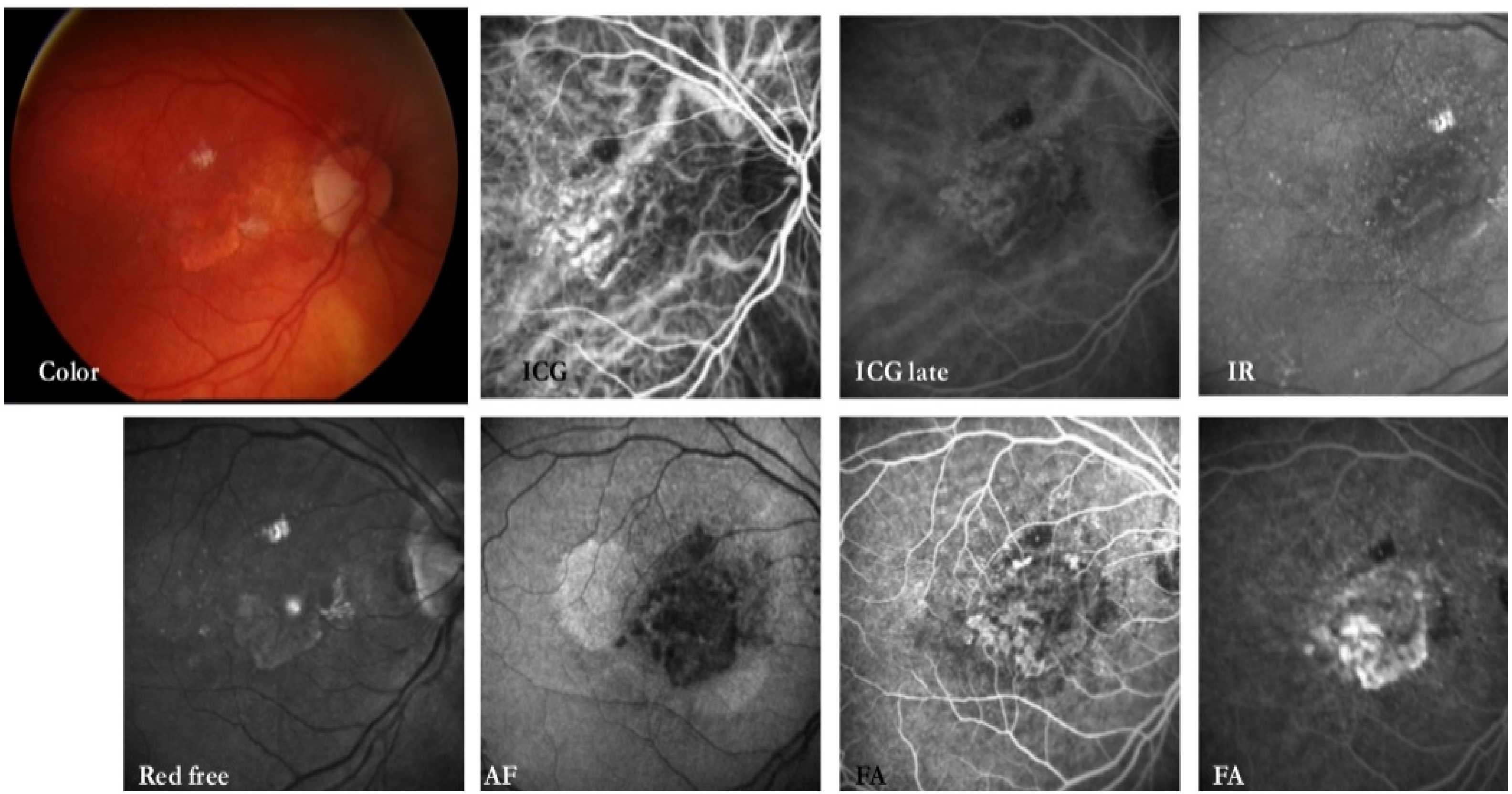


Autofluorescence eye fundus image

Infra-red eye fundus image

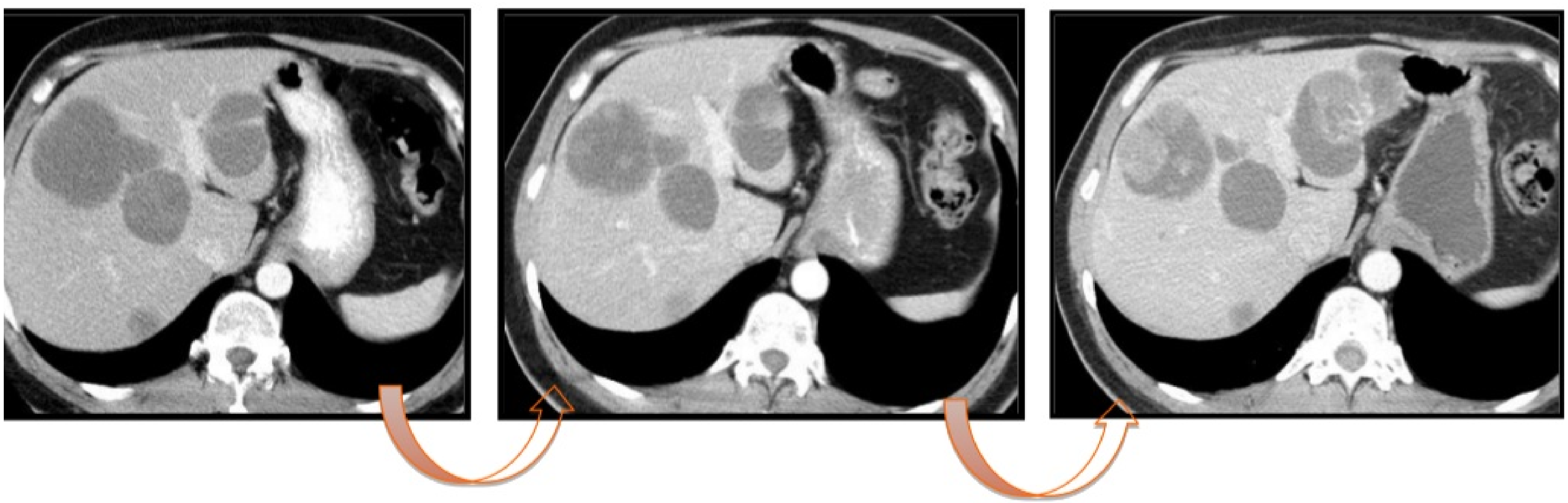
Etienne Decencière, CMM, Ecole des Mines de Paris

Inter-modality registration for eye fundus imaging



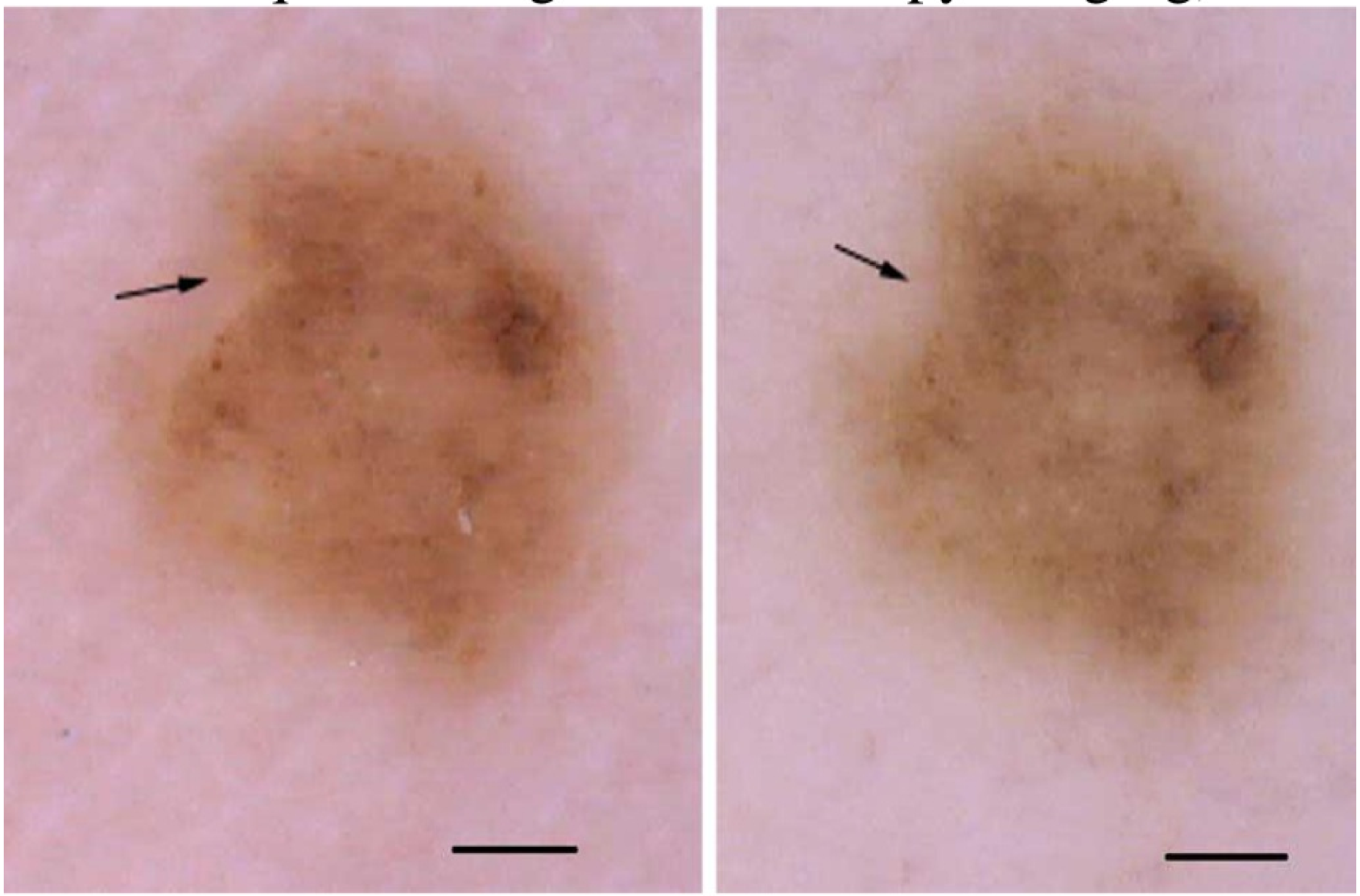
Registration is a generic issue in medical images :
- either between modalities (inter) or
- or within the same modality between time steps (or viewpoint) (intra).

Intra-modality registration: tumor and necrosis evolution



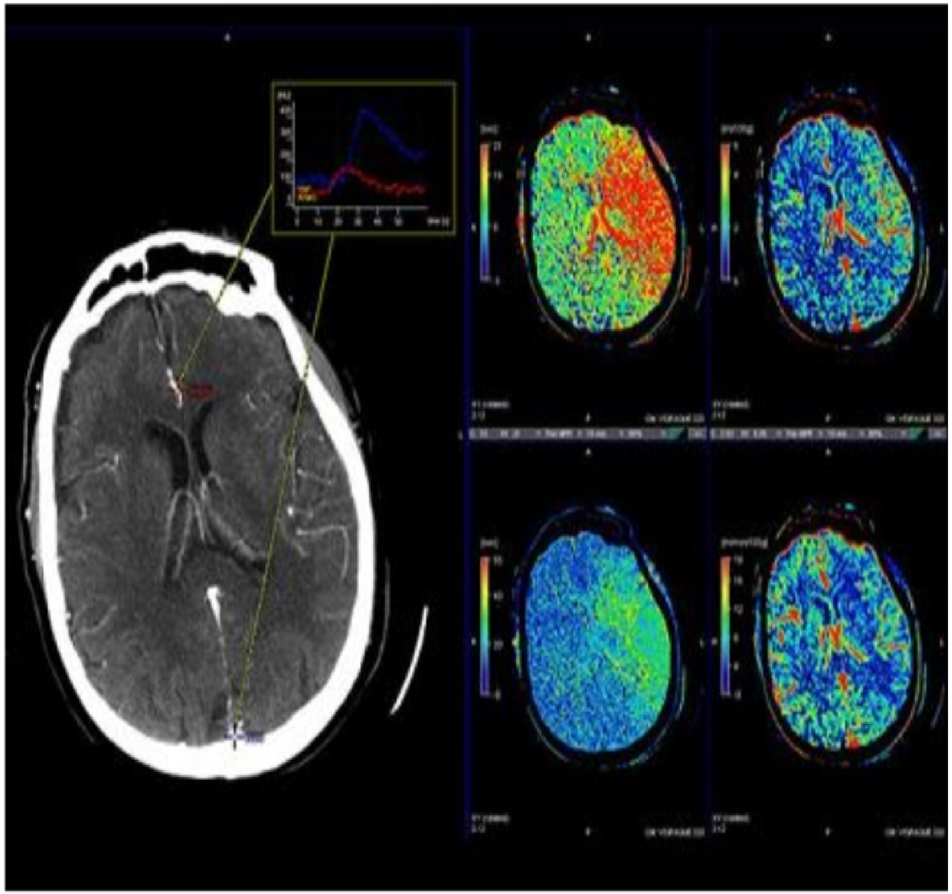
Source: Institut Gustave Roussy

Superficial spreading melanoma (6 weeks sequential digital dermoscopy imaging)

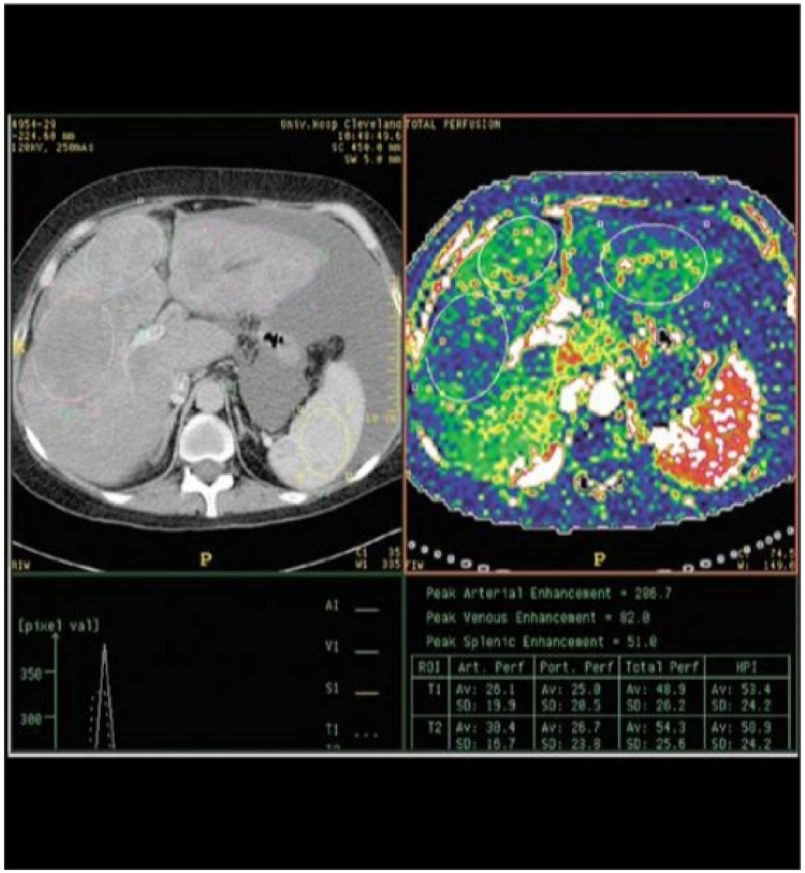


Altamura et al. Arch. Dermatol. (2008) 144

Dynamics analysis - perfusion



CT perfusion scan (source: Visage Imaging)

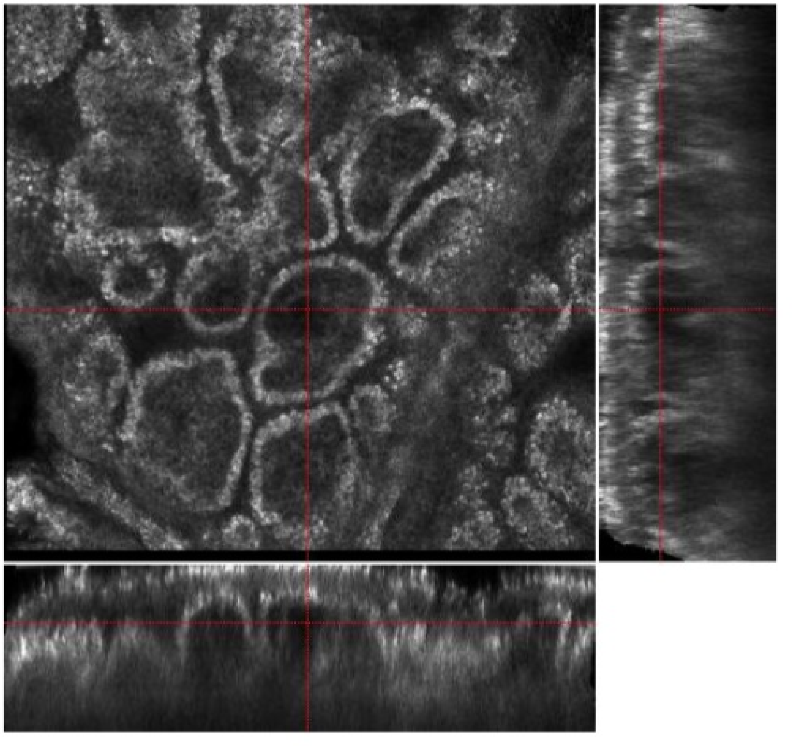


CT perfusion scan (source: Philips)

https://visageimaging.com/

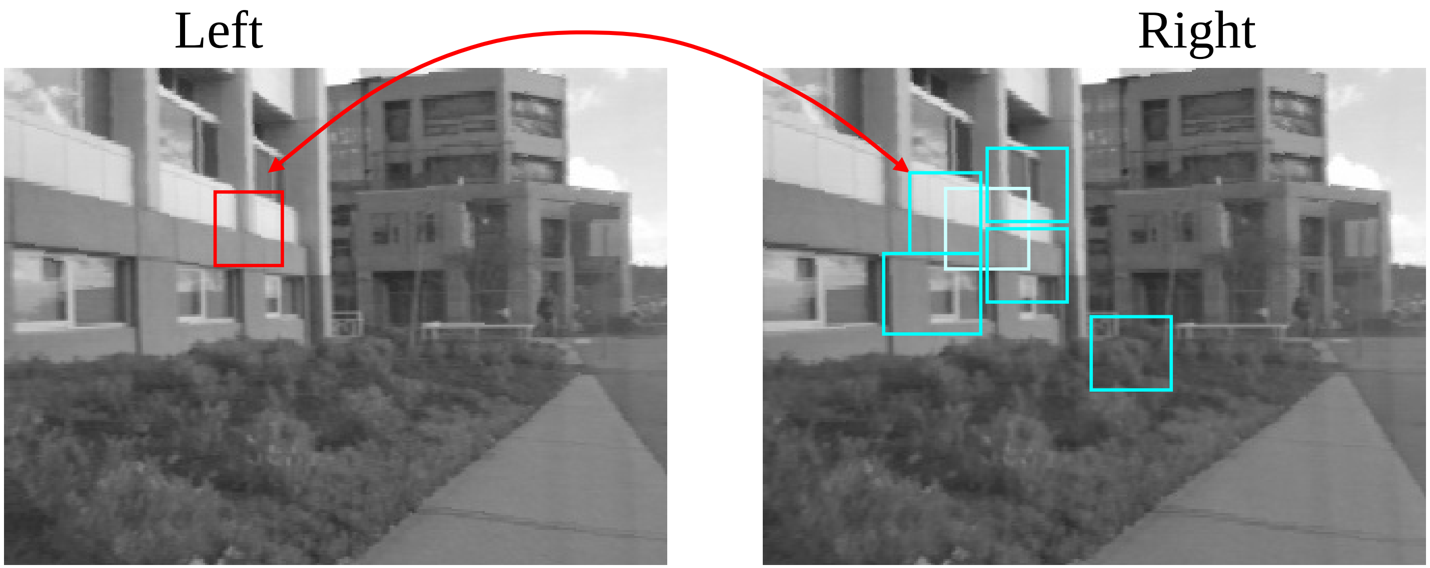
Registration may also be used for 3D reconstruction of organism from 2D slices or voxels data.

Registration of 3D stacks



Confocal microscopy stack of the skin

Take Home Messages



Applications *(Computer Vision vs. Bio-Medical Imaging)* are different but techniques are similar : **how to match patterns between a series of image (at least 2)** ?

Differences : the constraints / *a priori* knowledge (like anatomy)

You really learn and understand by teaching or implementing :-)

MASTER
BME

Video Analysis

Université de Paris
Nicolas Loménie

Motion

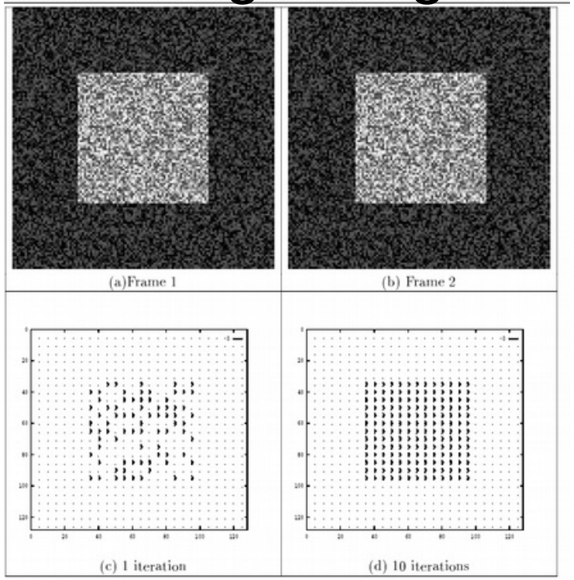
OPTICAL FLOW

Objectives :

- Understand the concept of motion field and optical flow;

- Knowledge of the brightness constancy equation;

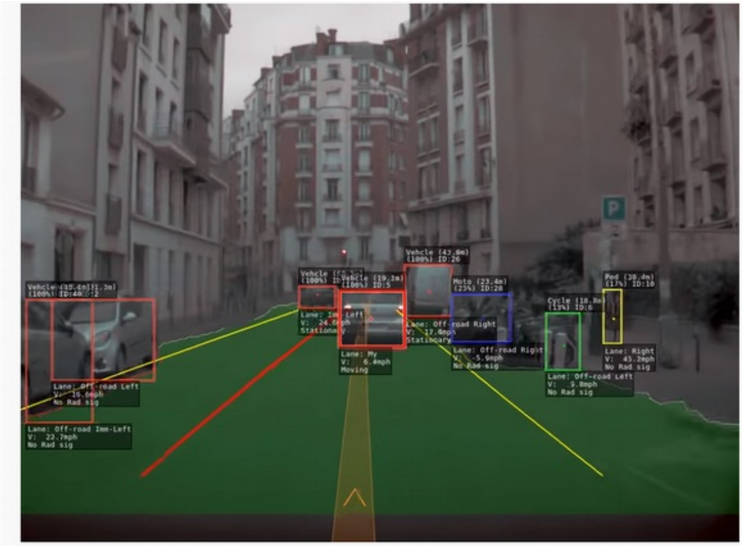
- Being able to understand/implement methodologies/algorithm for optical flow computation;



Results for Horn and Schunck algorithm for displacement of 1 pixel and $\lambda = 4$.

Paris streets

in the eyes of Tesla Autopilot



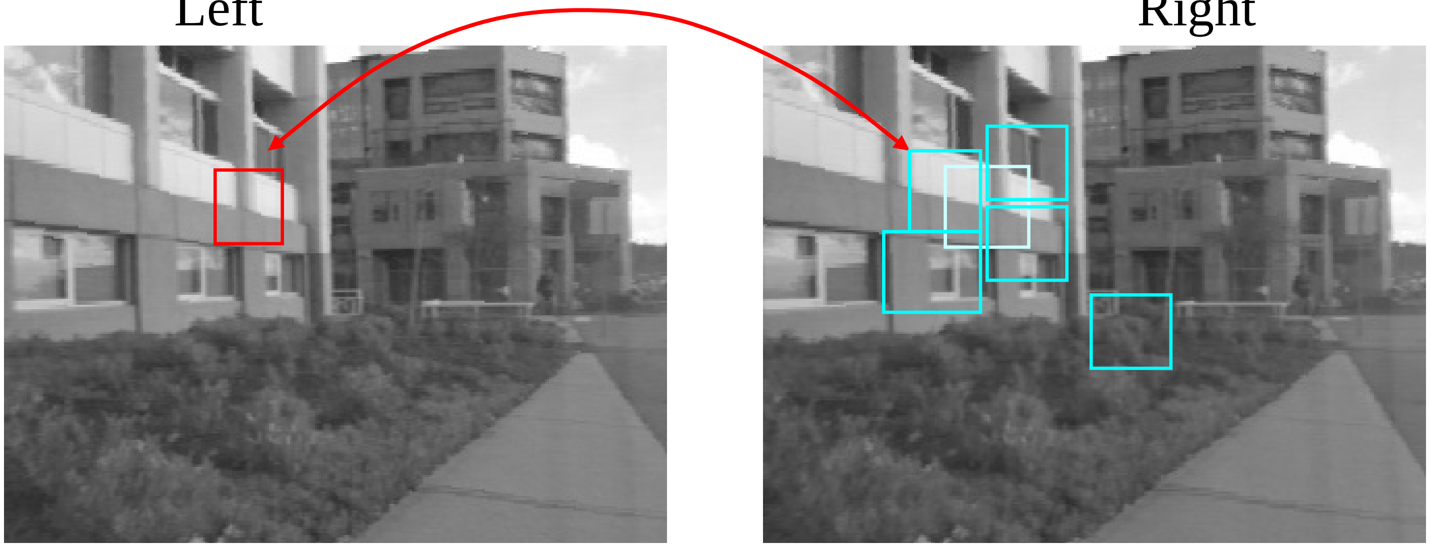
https://www.youtube.com/watch?v=_1MHGUC_BzQ

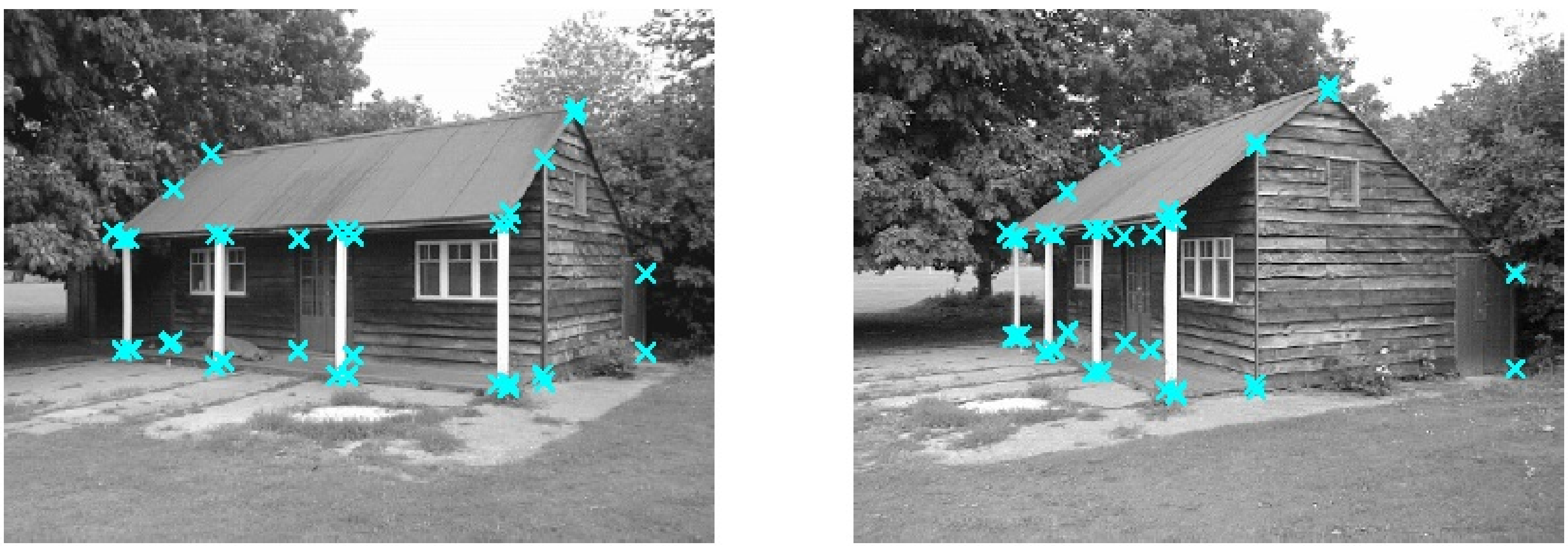
The matching issue: block matching strategy

Greedy search ?

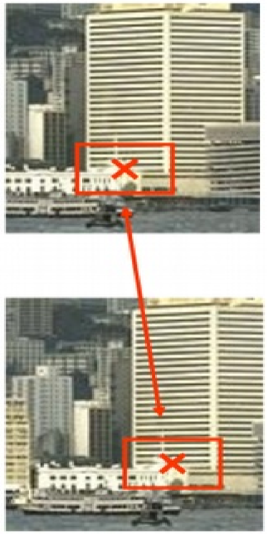
Left

Right





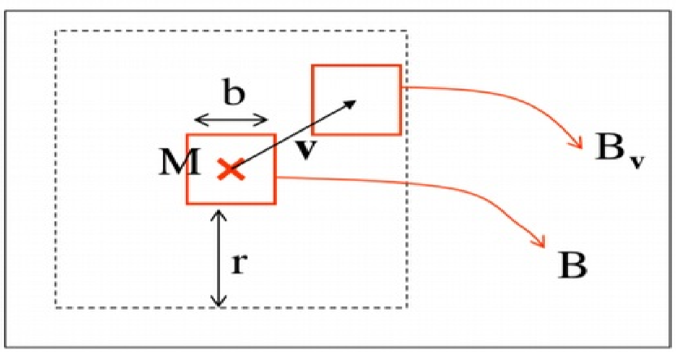
Block-Matching principle between 2 images I(t) and I(t+1), or I_{left} and I_{right} (stereo)
Or any pair of image I_1 and I_2 (registration in general or indexation)



Let us describe a motion estimation algorithm and its possible improvement strategies.

Block matching is extensively used on video compression (MPEG, H-263).

Basic idea: for computing the pixel displacement
- a block is set around the pixel,
- we search the best correspondence of this block over the target image. In the case of equal correspondence, the closest block is chosen.



B: bloc around point M.
B_v: B translated by **v**.
b: block size
r: « radius » of searching zone.

<u>Exhaustif block matching algorithm</u>. For computing motion vector of point M:
- For each possible vector v compute the correspondence error E(v) between B and B_v:

$$E(\vec{v}) = \sum_{M \in B} \left| I(M,t) - I(M+\vec{v},t+1) \right|$$

- Among all v that minimize E(v), take the one with smallest norm

Algo CORR_MATCHING

INPUT :

- Pair of images I_l and I_r
- Let p_l and p_r be the pixels in images I_l and I_r respectively. 2W+1 the size in pixels of the correlation window/matrix. $R(p_l)$ the search region in image I_l corresponding to p_l.
- Let $\Psi(u,v)$ be a function of two pixel values.

For every pixel $p_l=[i,j]^T$ in image I_l :

- For every displacement $d=[d_1,d_2]^T \in R(p_l)$ compute :

$$c(d)=\sum_{k=-W}^{W}\sum_{k'=-W}^{W}\Psi\left(I_l(i+k,j+k'),I_r(i+k-d_1,j+k'-d_2)\right)$$

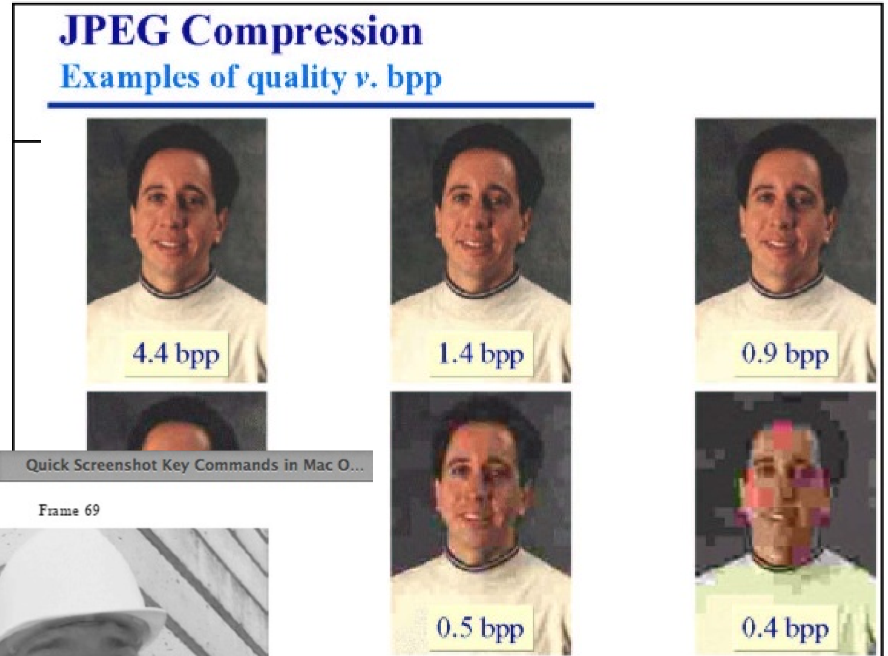
- The flow/displacement of p_l is the vector $\bar{d}=\left[\bar{d}_1,\bar{d}_2\right]^T$ maximizing *c(d)* over *R(p_l)* :

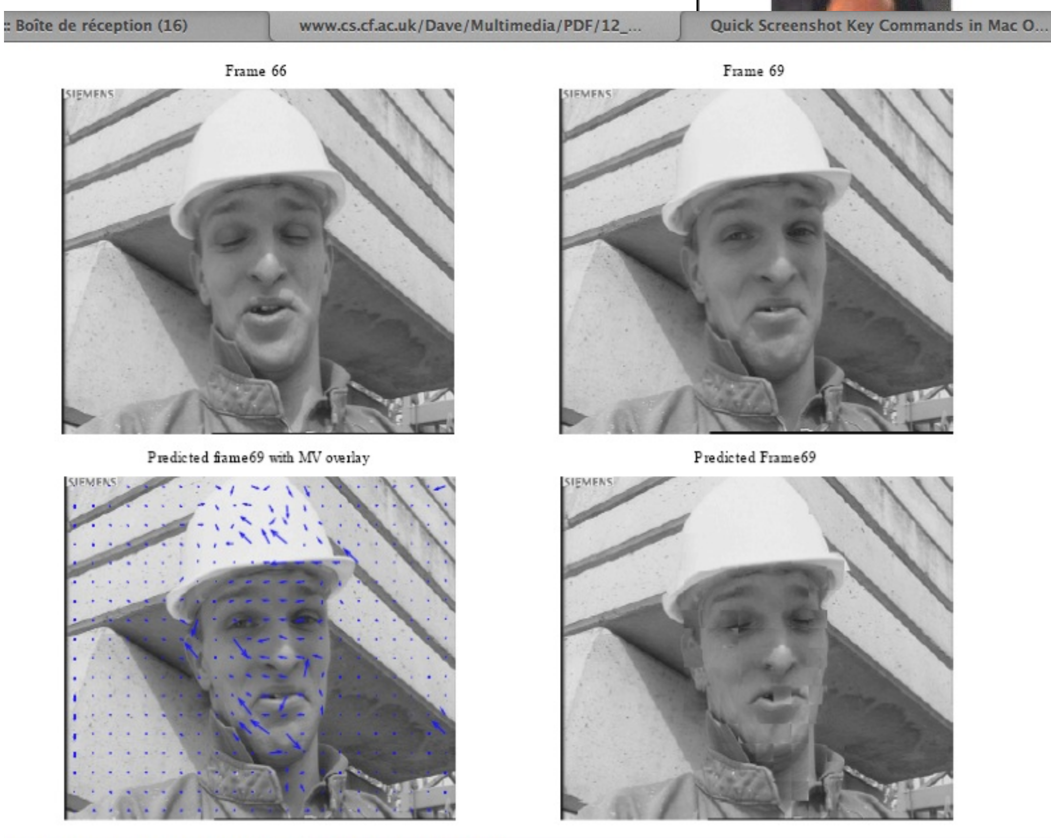
$$\bar{d}=Arg\max_{d\in R}\{c(d)\}$$

OUTPUT : set of displacements for every pixel in I_l.

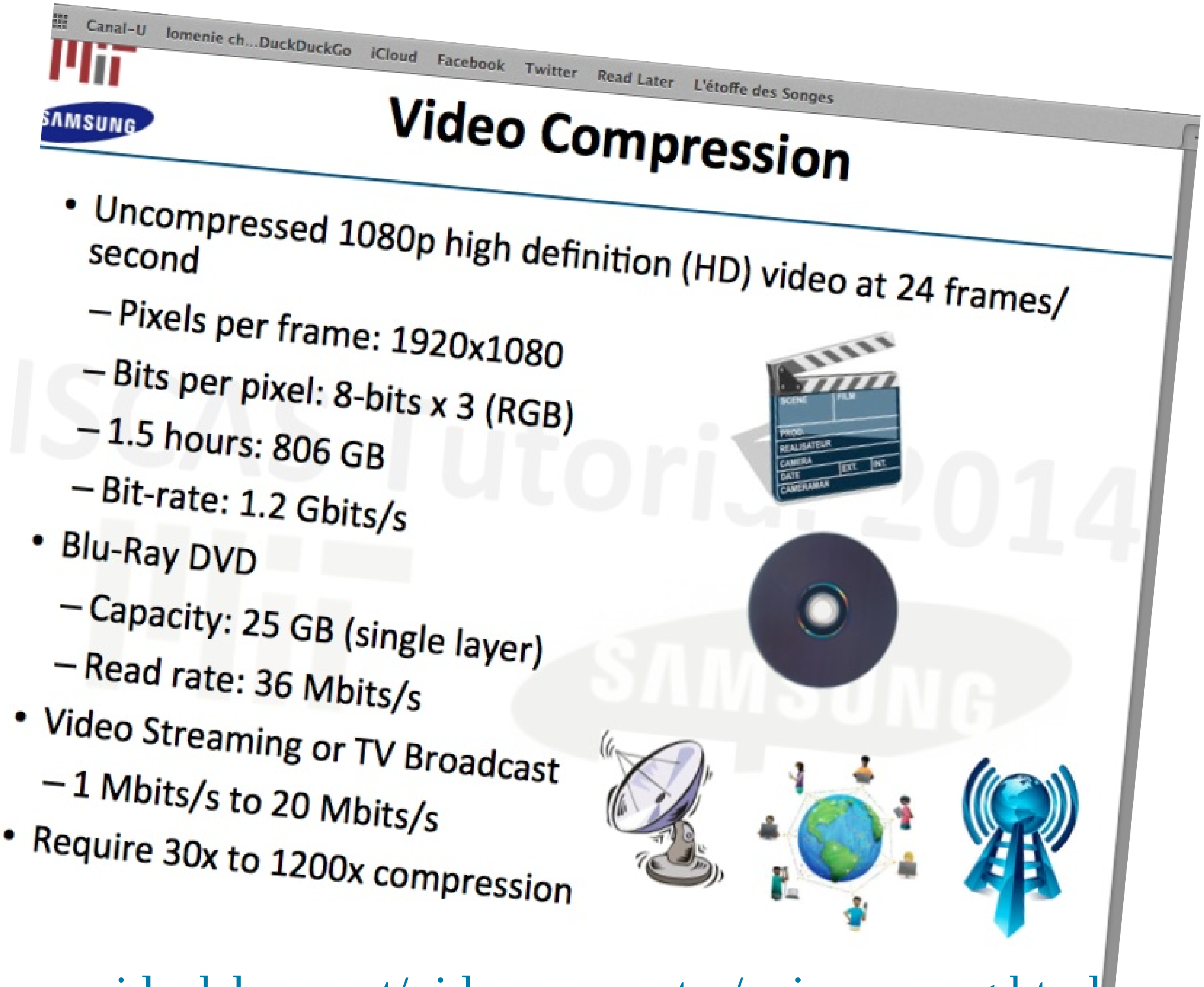
$$C(x,y,d)=\frac{\sum_{i,j}\left[\left(I_1(x+i,y+j)-\overline{I_1(x,y)}\right)-\left(I_2(x+i+d,y+j)-\overline{I_2(x+d,y)}\right)\right]^2}{\sqrt{\sum_{i,j}\left(I_1(x+i,y+j)-\overline{I_1(x,y)}\right)^2 * \sum_{i,j}\left(I_2(x+i+d,y+j)-\overline{I_2(x+d,y)}\right)^2}}$$

e.g. if only d on x axis (human stereo), normalized cross-correlation



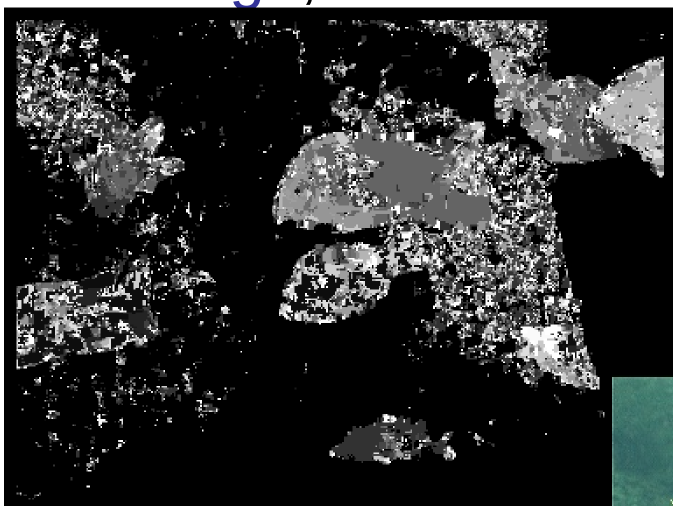


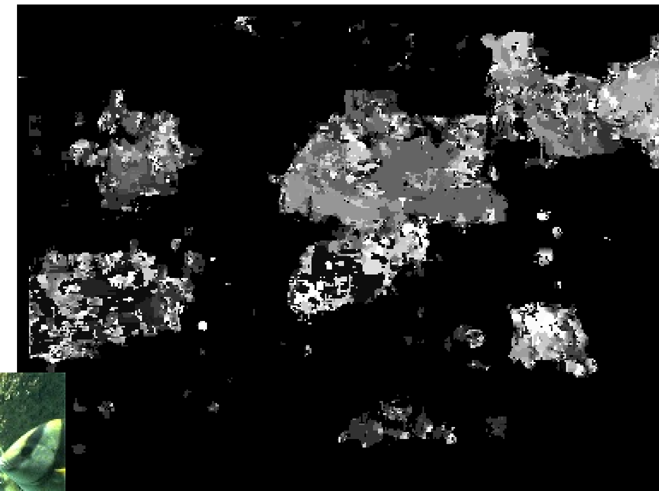
MPEG Compression

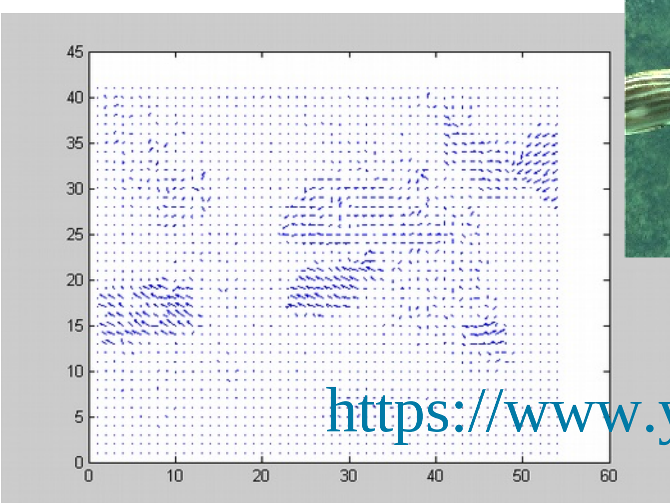


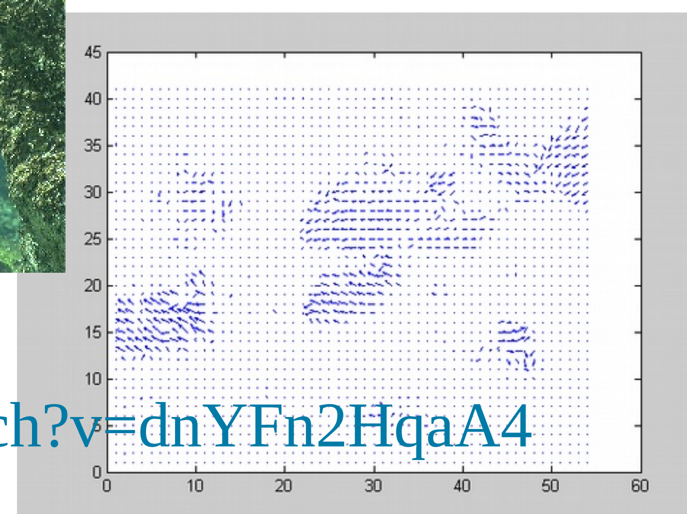
https://www.idealshare.net/video-converter/avi-vs-mpeg.html

In Videos, due to the t'-t << epsilon hypothesis : the matching issue is from another nature. It becomes the apparent motion estimation of radiometric patterns in the image, what is called the optical flow.



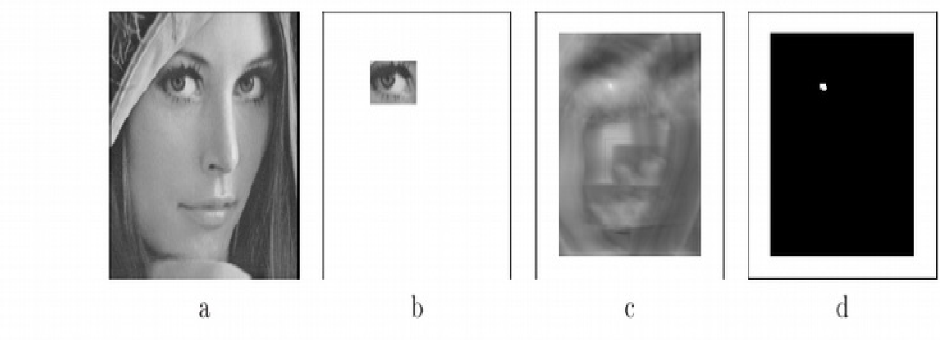


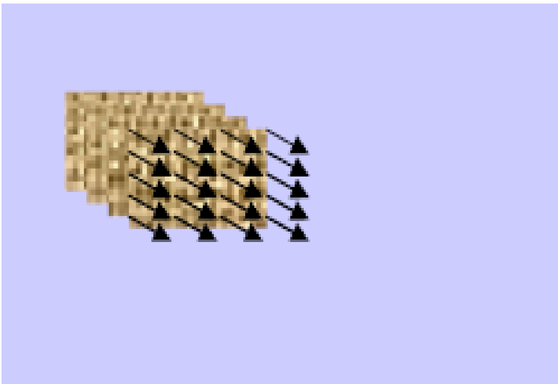


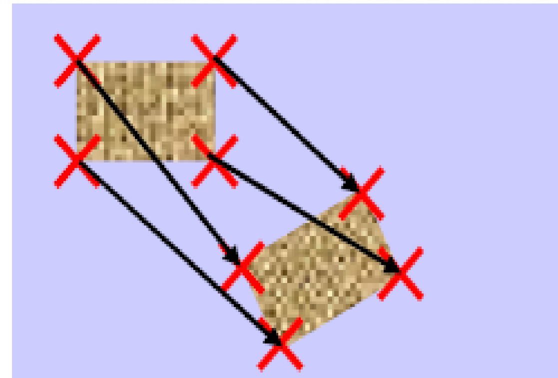


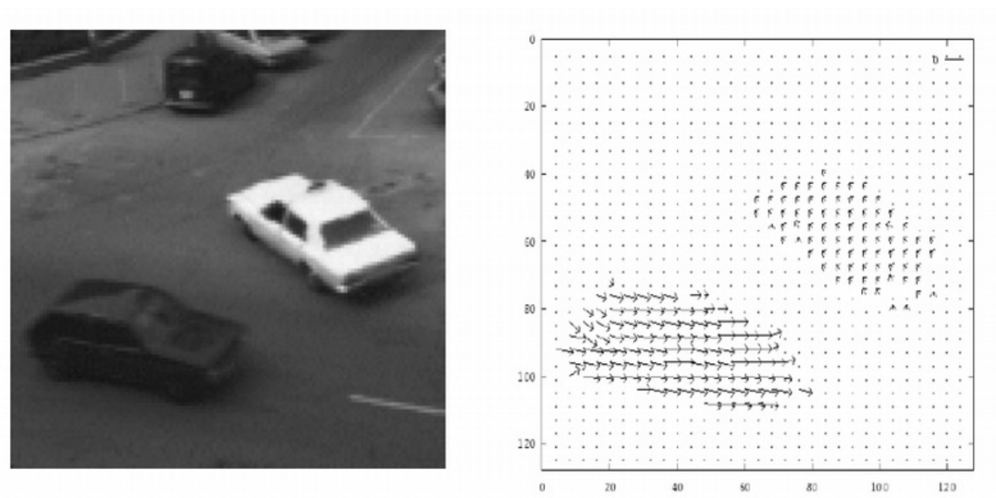
https://www.youtube.com/watch?v=dnYFn2HqaA4

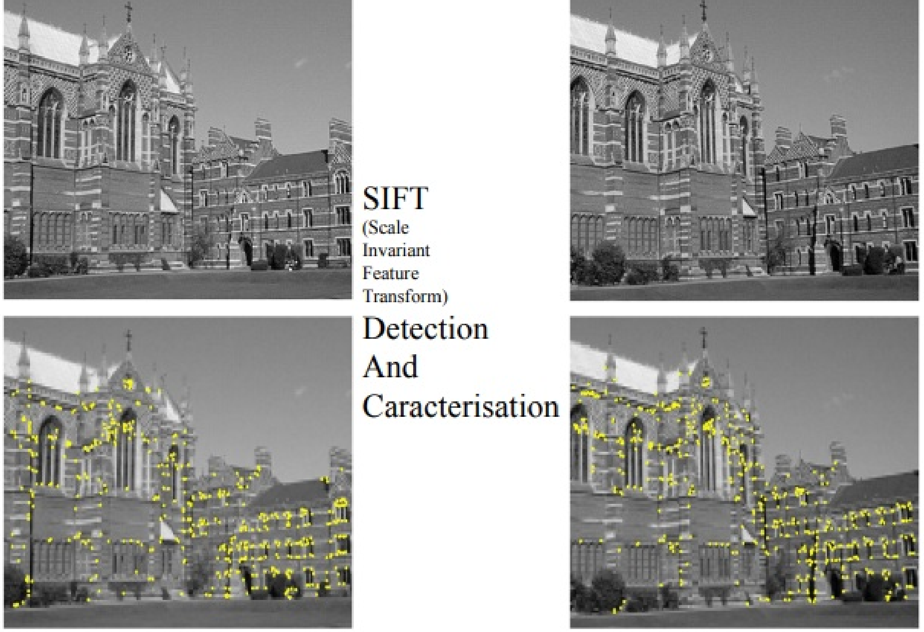
Primate biological vision ?



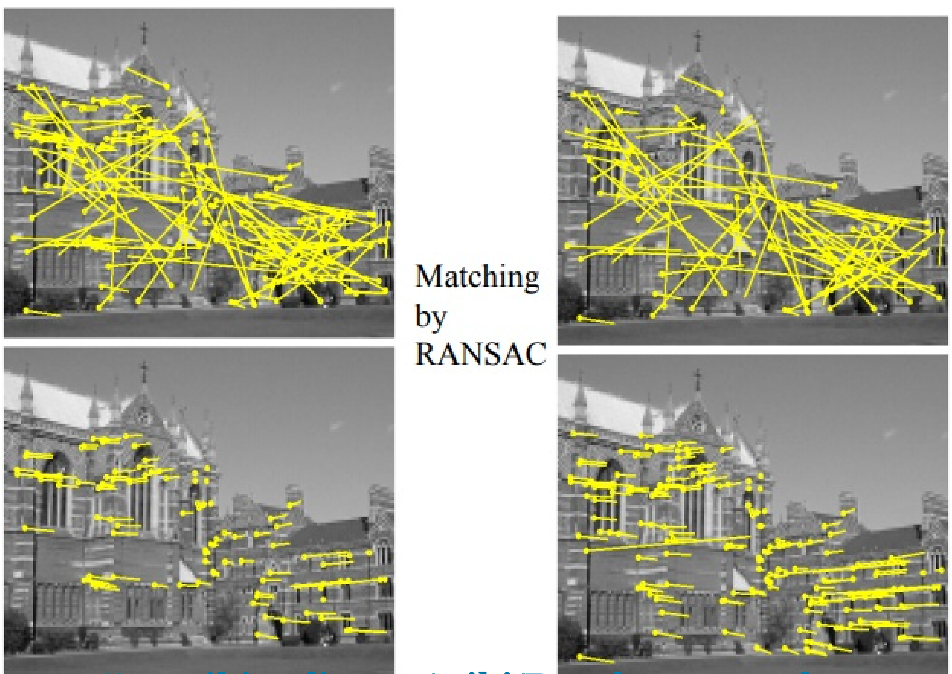




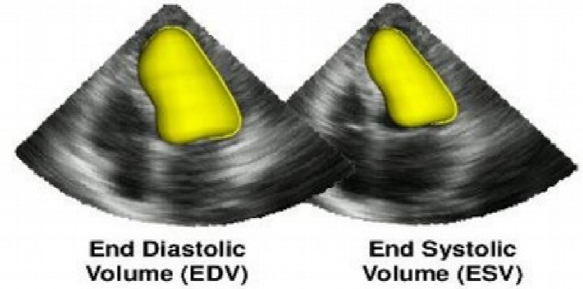




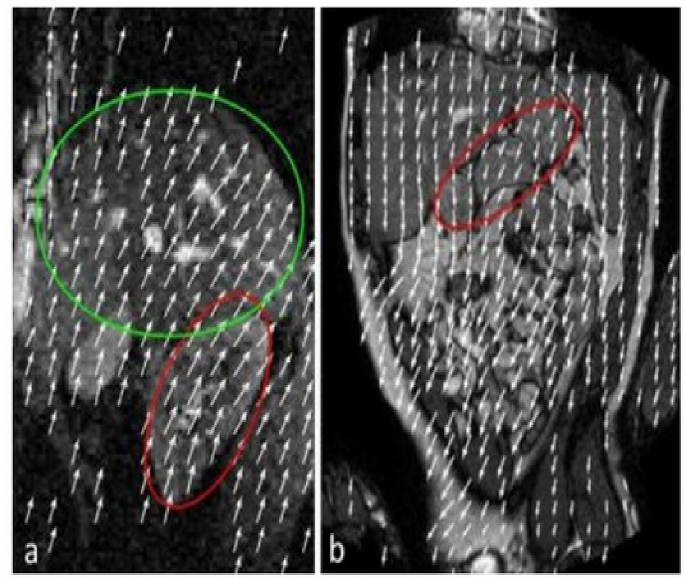
https://image-matching-workshop.github.io/



https://en.wikipedia.org/wiki/Random_sample_consensus

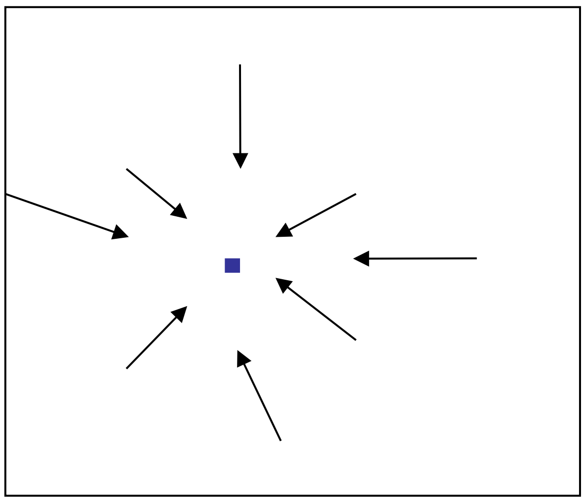


RealTITracker A toolbox for real-time 2D/3D optical flow based medical image registration

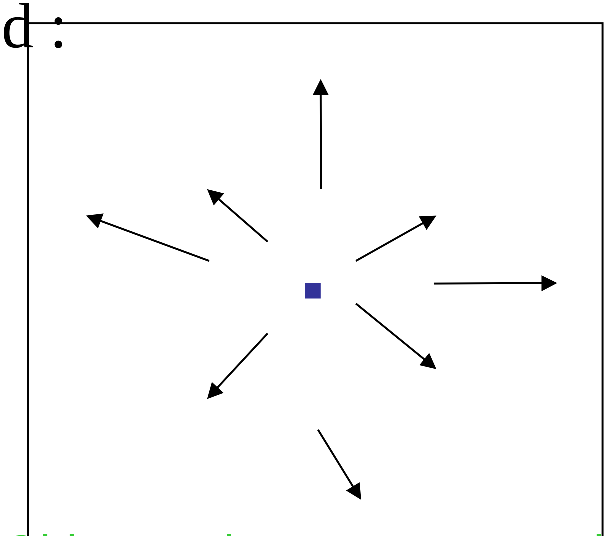


http://bsenneville.free.fr/RealTITracker/

If an algorithm is available to compute the motion in digital images, what should we observe/compute if we observe a pure translation moving object as a resulting vector field :

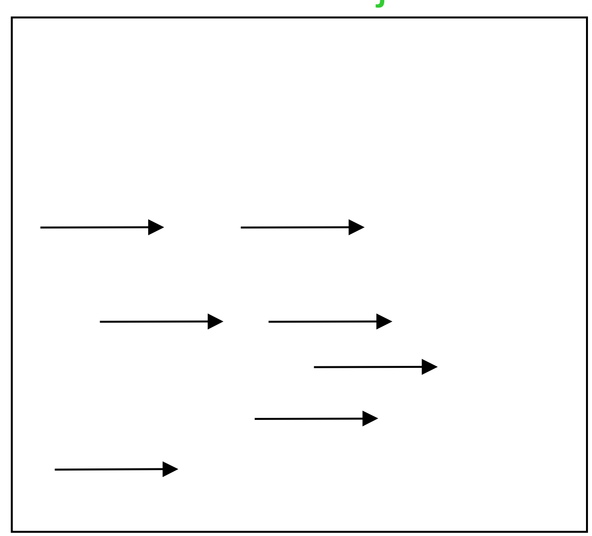


Object and camera are moving away



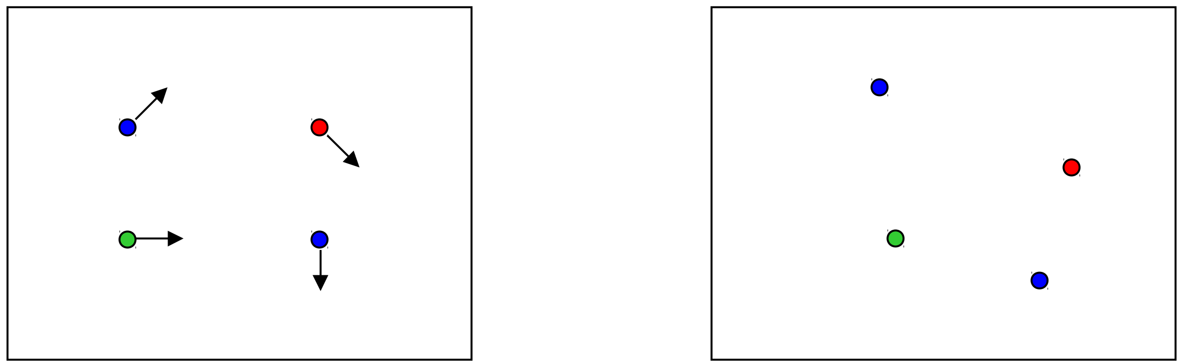
Object and camera are moving closer

Parallel motion of objects and camera



Problem definition: optical flow, how to compute it ?

$$\frac{\Delta x}{\Delta t} \Leftrightarrow \frac{\Delta I(x,y)}{\Delta t}$$



H(x,y) E(x,y)

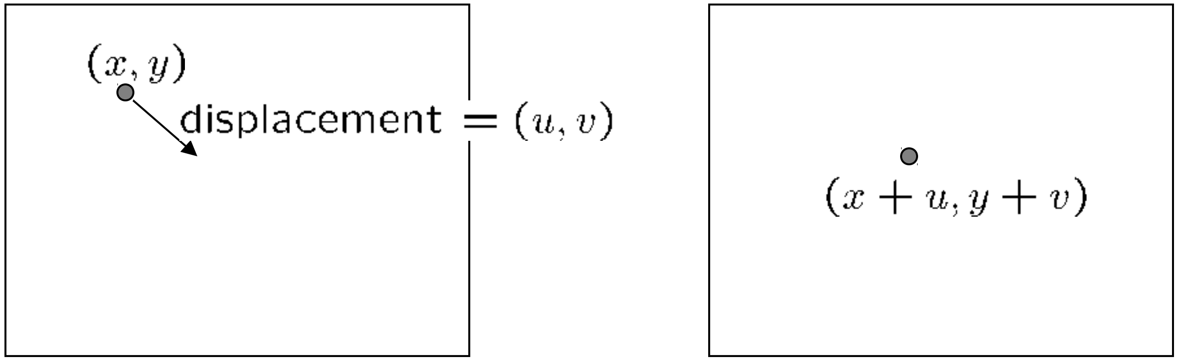
How to estimate pixel motion from image H to image E ?

- Solve pixel correspondence problem
  - given a pixel in H, look for nearby pixels of the same color in E

Key assumptions

- **color constancy:** a point in H looks the same in E
  - For grayscale images, this is **brightness constancy**
- **small motion**: points do not move very far away

Problem definition: optical flow : limits et constraints (brightness constancy)



H(x,y) E(x,y)

- Brightness constancy : $0 = E(x+u, y+v) - H(x,y)$
- Small displacements : u=Δx and v=Δy < 1 pixel

suppose we take the Taylor series expansion of E

$$E(x+\Delta x, y+\Delta y) = E(x,y) + \frac{\partial E}{\partial x}\Delta x + \frac{\partial E}{\partial y}\Delta y + O(\Delta x, \Delta y)$$

$$E(x+\Delta x, y+\Delta y) \approx E(x,y) + \frac{\partial E}{\partial x}\Delta x + \frac{\partial E}{\partial y}\Delta y$$

Combining these two equations

$$0 = E(x + \Delta x, y + \Delta y) - H(x, y)$$

$$0 \approx E(x, y) + E_x \Delta x + E_y \Delta y - H(x, y) \text{ avec } \mathrm{E}_{\mathrm{x}} = \frac{\partial E}{\partial x}$$

$$0 \approx (E(x, y) - H(x, y)) + E_x \Delta x + E_y \Delta y$$

$$0 \approx \Delta E + E_x \Delta x + E_y \Delta y$$

$$0 \approx \Delta E + \overrightarrow{\nabla E} \cdot \begin{bmatrix} \Delta x & \Delta y \end{bmatrix}$$

$$0 \approx \frac{\Delta E}{\Delta t} + \overrightarrow{\nabla E} \cdot \begin{bmatrix} \frac{\Delta x}{\Delta t} & \frac{\Delta y}{\Delta t} \end{bmatrix}$$

In the limit as u and v go to zero, this becomes exact :

$$E_t + \overrightarrow{\nabla E} \cdot \begin{bmatrix} \frac{dx}{dt} & \frac{dy}{dt} \end{bmatrix} = 0$$

And so the following fundamental equation :

$$\frac{dE(x, y, t)}{dt} = 0$$

As x(t) and y(t)

$$\frac{dE(x(t),y(t),t)}{dt} = \frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0$$

It expresses the common sense : $\frac{\Delta x}{\Delta t} \Leftrightarrow \frac{\Delta I(x,y)}{\Delta t}$

Basic Image Processing lectures :

Contour maps :

Gradient Vector at P(x,y)

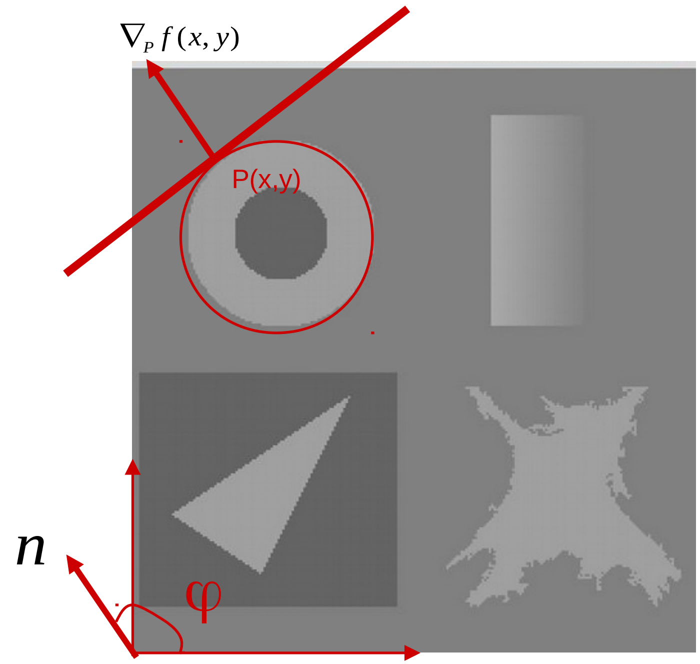
$$\vec{\nabla}_P f(x,y) = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix}$$

$$\varphi = Arc\tan\left(\frac{\partial f}{\partial y} \Big/ \frac{\partial f}{\partial x} \right)$$

Contour orientation : $\theta = \frac{\pi}{2} + \varphi$

N = normal vector to the level set

f(x,y) = f(x_P,y_P)=cst,



Where

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial x} \\ \frac{\partial E}{\partial y} \end{bmatrix}$$ (Frame spatial gradient)

$$v = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}$$ (optical flow)

and

$$E_t = \frac{\partial E}{\partial t}$$ (derivative across frames)

Brightness Constancy Equation

E(x,y,t) being the image brightness and **v the motion field**, we write :

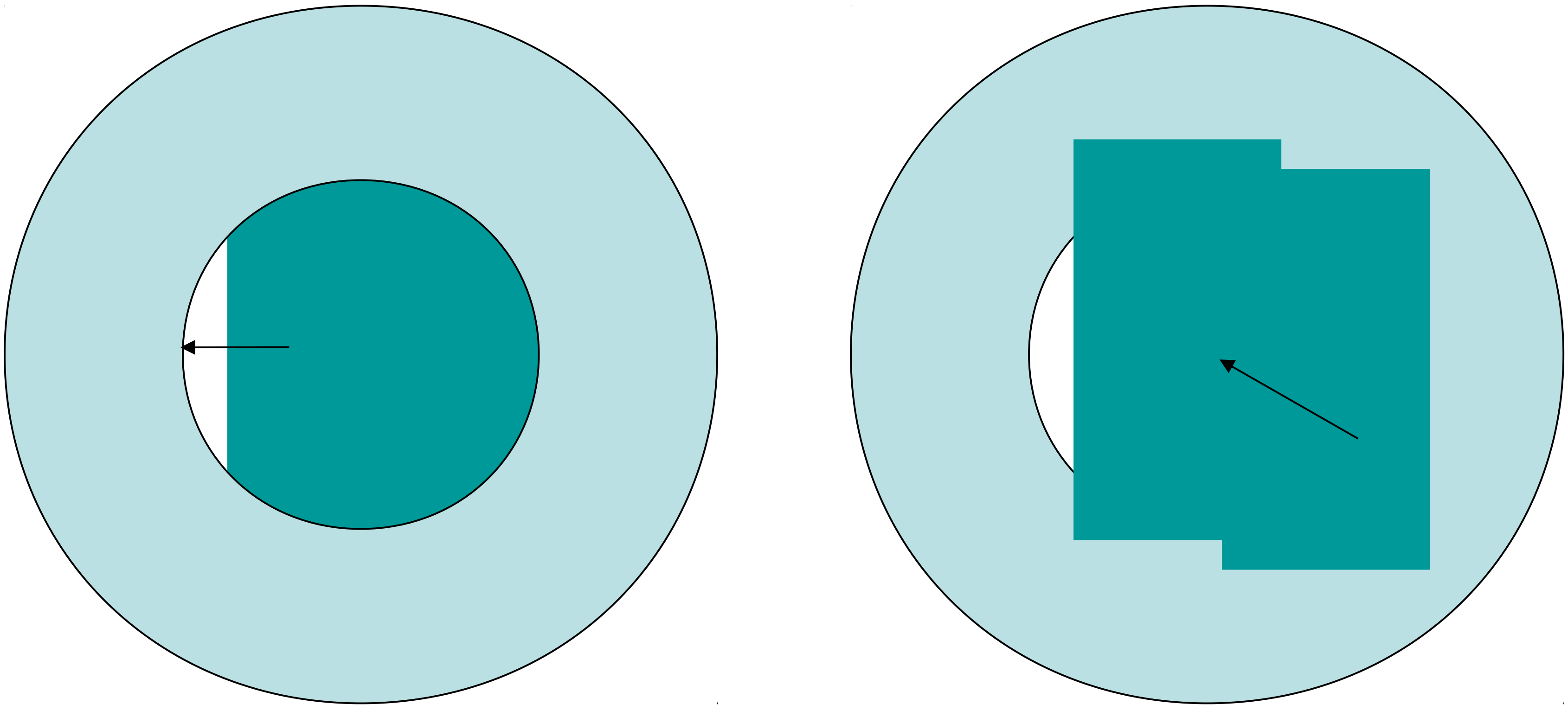
$$(\nabla E)^T v + E_t = 0$$

where E_t is the temporal partial derivative.

As t'-t<<epsilon, we can compute/measure ∇E and E_t (image processing), hence **v** is not far. So what ?

How an algorithm can estimate the motion field v from this equation ?

The aperture problem



The Image Brightness Constancy Assumption only provides the OF component in the direction of the spatial image gradient

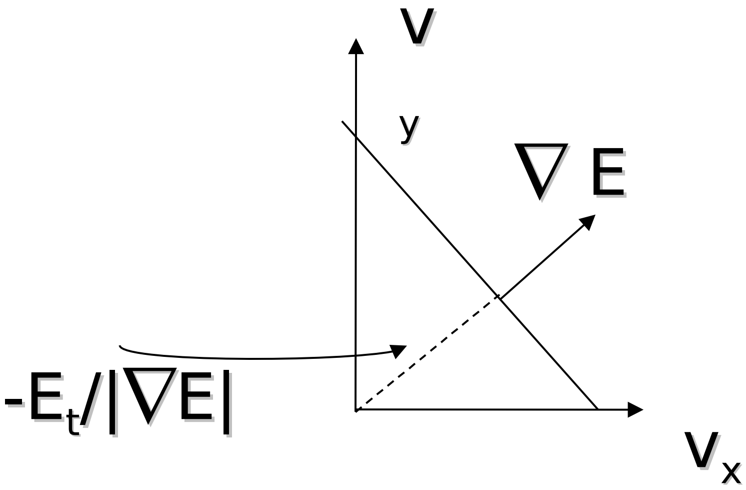
Intuitively, what does this constraint mean?

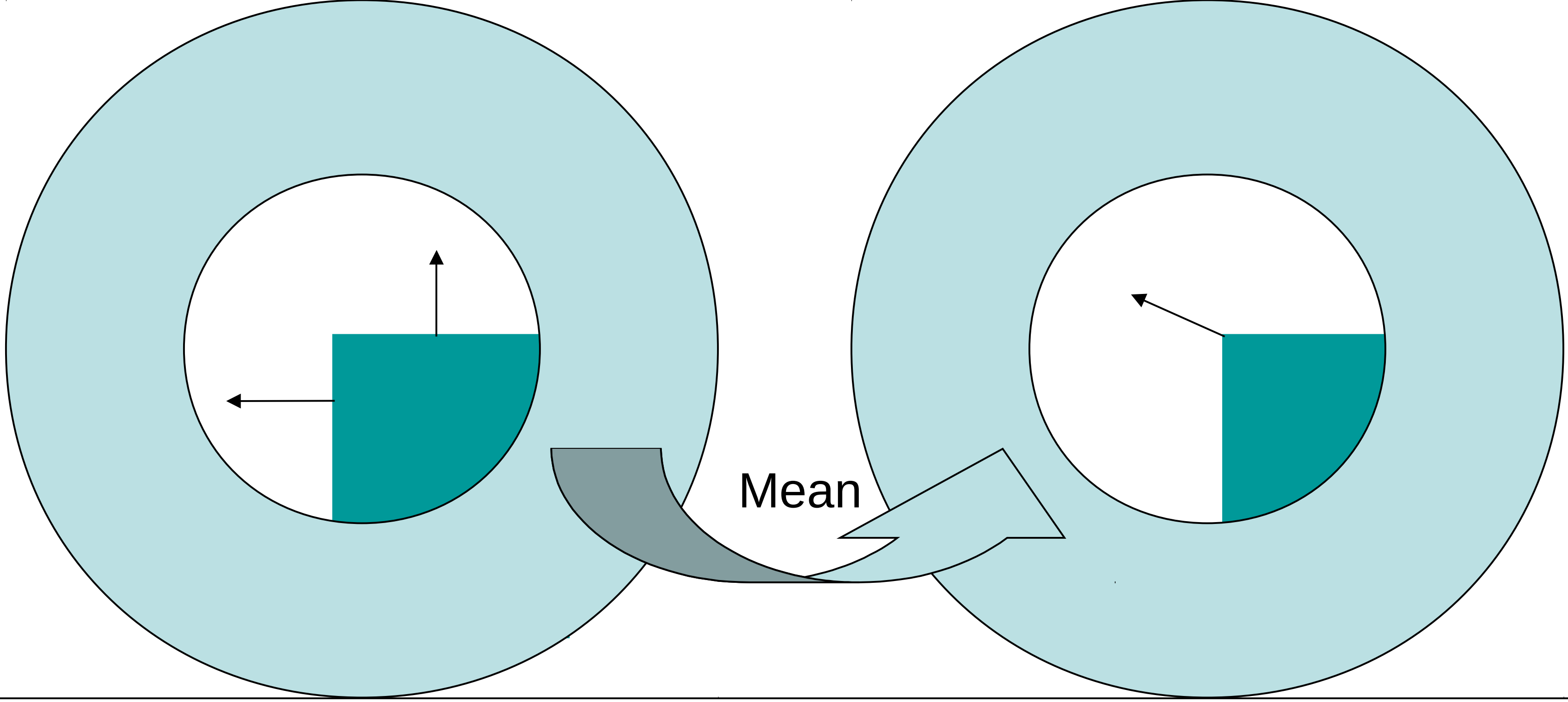
$$\boxed{(\nabla E)^T . v + E_t = 0}$$

$$\frac{(\nabla E)^T v}{\|\nabla E\|} = - \frac{E_t}{\|\nabla E\|} = v_n$$

- The component of the flow in the gradient direction is determined
- The component of the flow parallel to an edge is unknown

$$\frac{\partial E}{\partial x} \rightarrow \frac{\Delta E}{\Delta x}$$





Optical Flow

The optical flow is a motion field satisfying the brightness constancy equation:

$$(\nabla E)^T.v + E_t = 0$$

Motion field Estimation

From an image sequence

Two kind of algorithmic approaches

- Dense matching techniques: **differential techniques (like PDE) :**
  -> optical flow methods

Genuine optical flow

- Sparse matching techniques :
  -> tracking methods

Can do without "brightness constancy" by using geometric constraints

Differential Technique: Optical Flow direct computation by MSE regression

For every pixel $\mathbf{p}_i$ inside a small patch **Q** of size ***NxN*** *(e.g. 5x5)* we can write :

$$\left(\nabla E(p_i)\right)^T v(p_i) + E_t(p_i) = 0$$

$$\text{and} \quad \forall i, \quad v(p_i) = v_Q$$

$$\text{hence} \quad \forall p_i \in Q, \quad \left(\nabla E(p_i)\right)^T v_Q + E_t(p_i) = 0$$

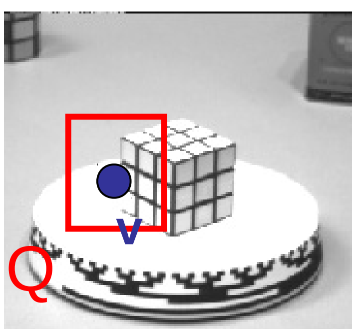
$$\Psi(v) = \sum_{p_i \in Q} \left| (\nabla E(p_i))^T v + E_t(p_i) \right|^2$$

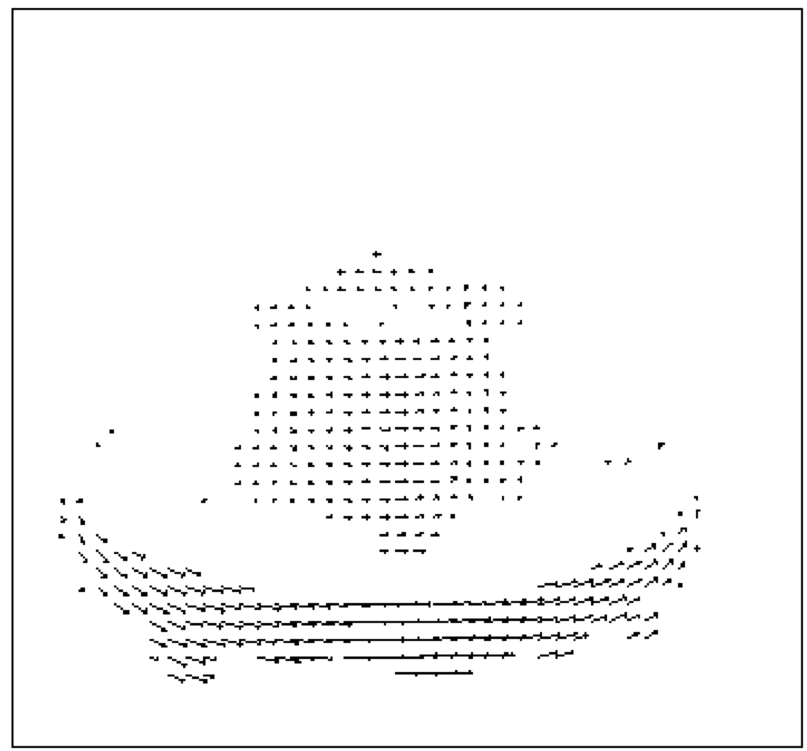
$$\Psi(v) = \sum_{p_i \in Q} \left| \left(\nabla E(p_i) \right)^T v + E_t(p_i) \right|^2$$

$$Av = b \qquad \begin{bmatrix} E_x(p_1) & E_y(p_1) \\ E_x(p_2) & E_y(p_2) \\ \vdots & \vdots \\ E_x(p_{25}) & E_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} E_t(p_1) \\ E_t(p_2) \\ \vdots \\ E_t(p_{25}) \end{bmatrix}$$

Whose solution is

$$v = \left(A^T A\right)^{-1} A^T b$$





Algorithm CONSTANT_FLOW

INPUT : A temporal sequence of n images E_1, E_2, ..., E_n. Let Q be a squared region of size NxN pixels (e.g. 5x5)

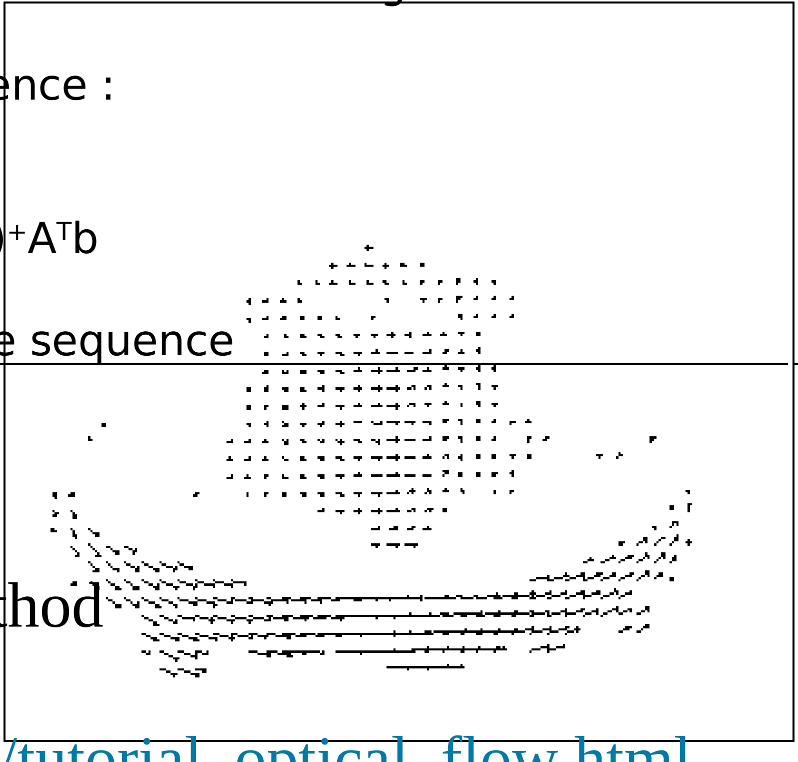
Filter every image of the sequence with a Gaussian filter of standard deviation σ_S (e.g. σ_S = 1,5 pixels) along each spatial dimension.

- Filter every image of the sequence with a Gaussian filter of standard deviation σ_S (e.g. σ_t = 1,5 frames) along the temporal dimension. If 2k+1 is the size of the temporal filter, you need not to process the k first and last images.

- For each pixel p of each image in the sequence :

    - Compute matrix A and vector b
    - Compute optical flow $v(p) = (A^TA)^+A^Tb$

OUTPUT : the optical flow of the entire image sequence



Key concept of the Lucas-Kanade method

Implemented in openCV for instance

https://docs.opencv.org/master/d4/dee/tutorial_optical_flow.html

This matrix is not invertible if

1. All spatial gradients within Q vanish
2. All spatial gradients within Q are parallel

$$A^T A = \begin{pmatrix} \sum E_x^2 & \sum E_x E_y \\ \sum E_x E_y & \sum E_y^2 \end{pmatrix}$$

Differential Technique: Optical Flow Iterative computation by PDE: Horn and Shunck

Algorithme HORN_FLOW

INPUT : Une séquence temporelle de n images f_1, f_2, ..., f_n.

1. Filtrer chaque image de la séquence avec un filtre de Dérivation le long de chaque dimension spatiale.
- Filtrer chaque image de la séquence le long de la dimension temporelle avec un filtre de Dérivation.

1. Pour chaque pixel p de chaque image de la séquence :
   - k=0
   - Initialise u^k et v^k à zéro

1. Pour chaque pixel p de chaque image de la séquence :
   - Jusqu'à ce qu'une mesure d'erreur soit satisfaite, faire :

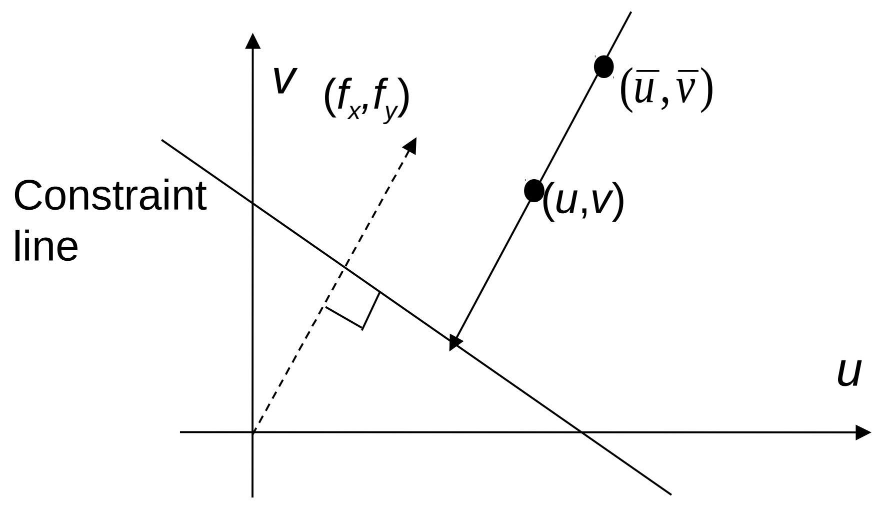
$$u^k = u^{k-1}_{moyen} - f_x \frac{P^{k-1}}{D}$$

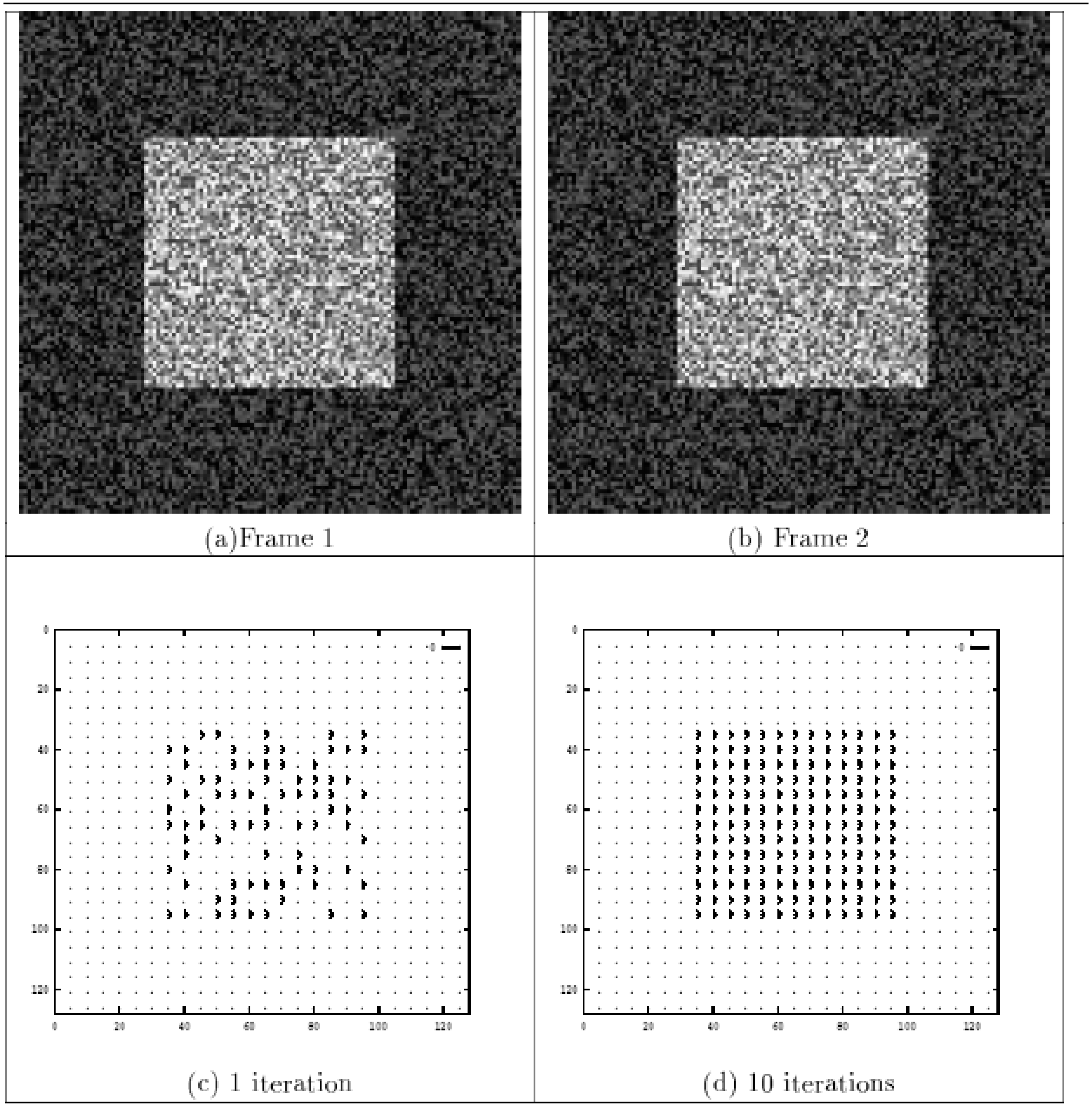
$$v^k = v^{k-1}_{moyen} - f_y \frac{P^{k-1}}{D}$$

OUTPUT : le flot optique de la séquence d'images

$$u = u_{moyen} - f_x \frac{f_x u_{moyen} + f_y v_{moyen} + f_t}{\lambda + f_x^{\,2} + f_y^{\,2}} = u_{moyen} - f_x \frac{P}{D}$$

$$v = v_{moyen} - f_y \frac{P}{D}$$





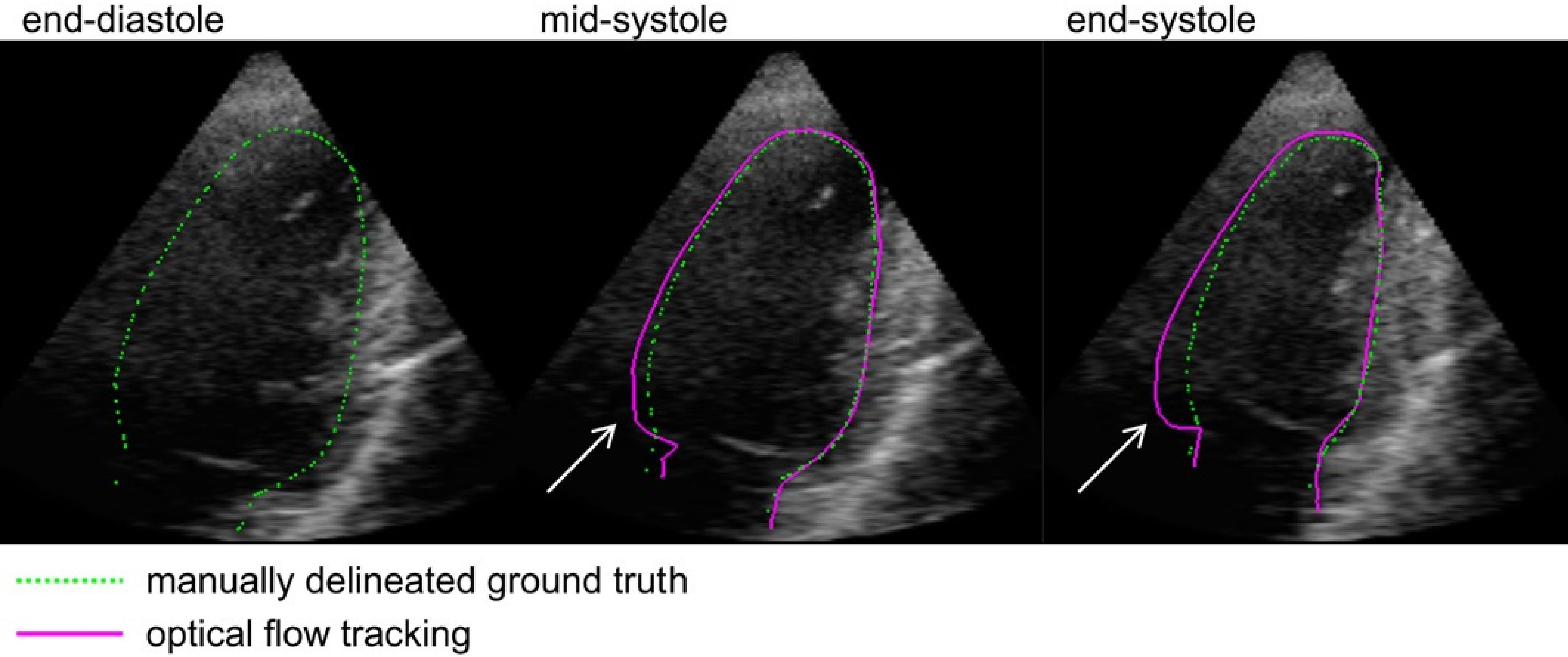
(a)Frame 1

(b) Frame 2

(c) 1 iteration

(d) 10 iterations

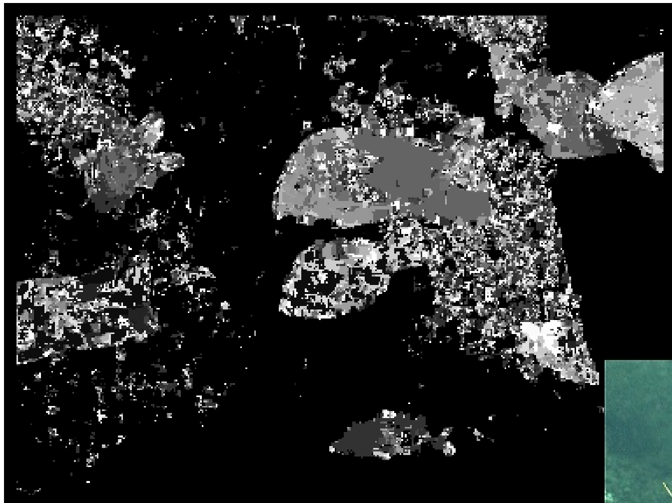
Results for Horn and Schunck algorithm for displacement of 1 pixel and $\lambda = 4$.

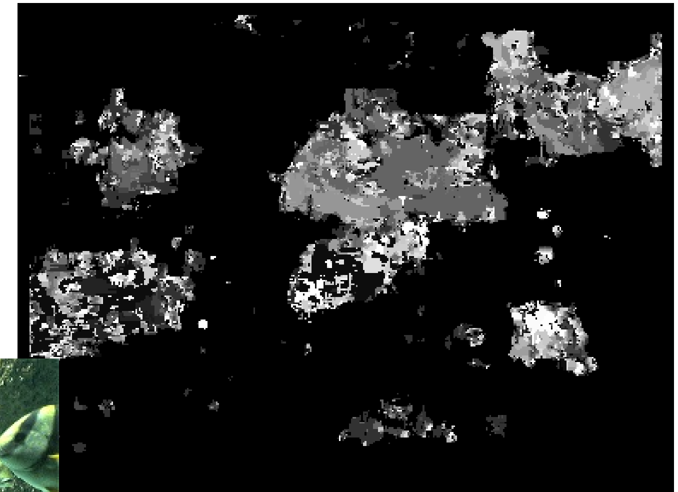


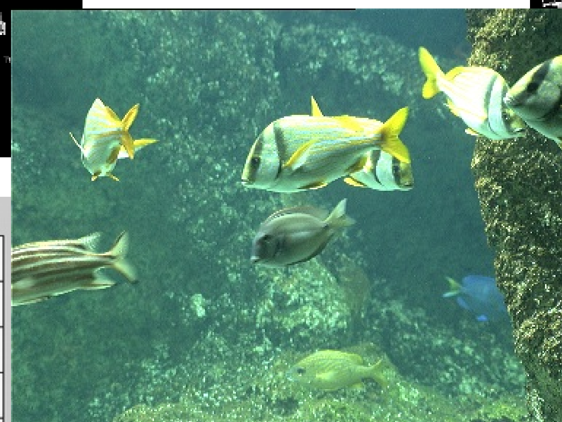
Example of echocardiographic sequence (two-chamber view is shown), with poorly visualized cardiac wall in the anterior segments. Green dotted line denotes the manually delineated ground truth. Magenta solid line denotes optical flow tracking. Misinterpretation of the anterior wall (arrow) may lead to considerable inaccuracies in quantification.

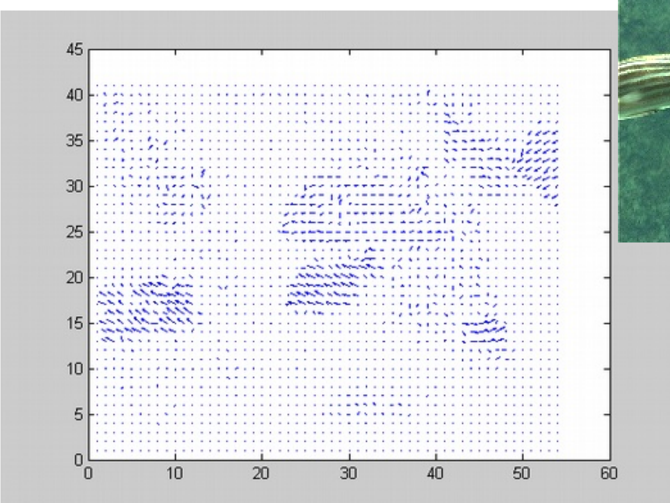
From Ultrasound in Med. & Biol., Vol. 37, No. 4, pp. 605–616, 2011

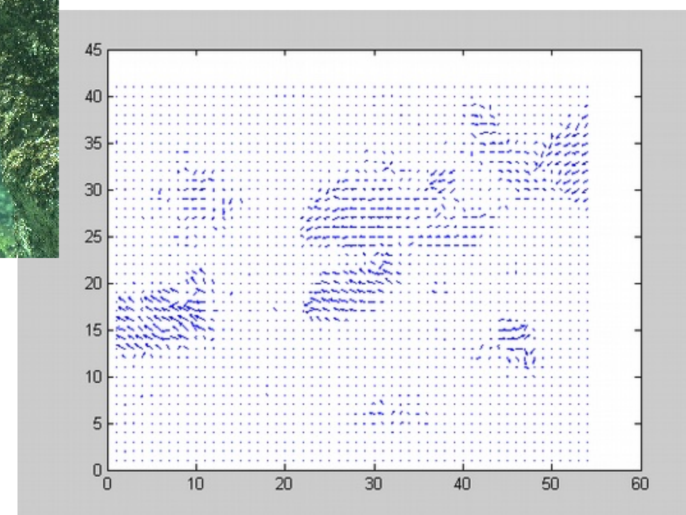
Once computed: motion-based segmentation











https://www.youtube.com/watch?v=h_Q7ADrrdAY

https://www.youtube.com/watch?v=Q5itLKscYTA&t=712s 11-14 min