MOTION ANALYSIS

computer vision

INTRODUCTION

- so far we have discussed way of inferring information from a single image
- from now on we will explicitly consider information we may extract from sets of images of the same scene
- this week (and the next) we will focus on *motion information*

INTRODUCTION

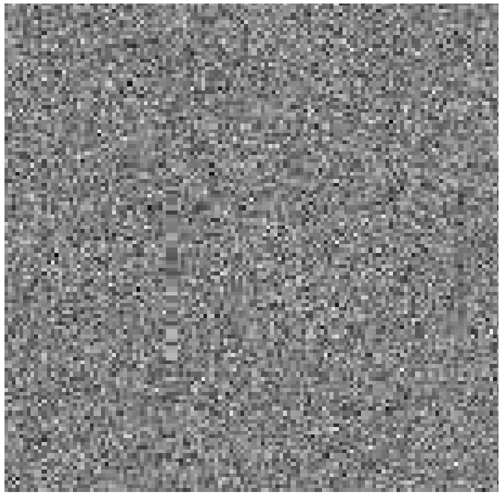
- We are observing a scene with **one** camera acquiring a set of images "close in time"

- **image sequence:** series of N images, or *frames*, acquired at discrete time instants

fixed time interval (small)

$$t_k = t_0 + k\Delta t$$

THE IMPORTANCE OF MOTION CUES



- the presence of a structure (square) is suggested to the visual system only by motion information: the changes we perceive from one frame to the next

INTRODUCTION

- An important **assumption** we will make from now on is the following:
  - *illumination does not vary in the observation interval*
- in this case the image changes from t_1 to t_2 are caused by the *relative motion* between scene and camera

THE PROBLEMS OF MOTION

- We may divide our problem in two more specific issues:
  - **correspondence:** which elements of a frame correspond to which elements of the next frame? (today)
  - **reconstruction:** given a number of corresponding elements what can we say about 3D motion (and 3D structure) of the scene? (later in the course, briefly)
- Since adjacent frames are similar, correspondence is rather easy, but reconstruction is hard!

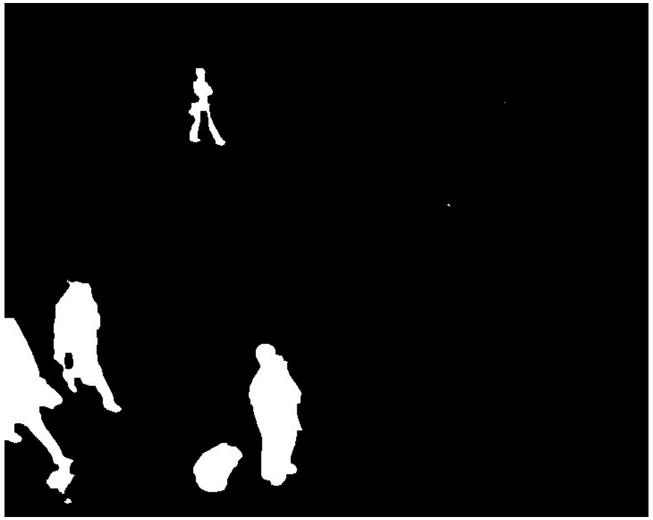
THE PROBLEMS OF MOTION (CONT'D)

- other problems related to motion analysis
  - **motion segmentation:** what regions of the image plane correspond to different moving objects? in the case the camera is still, motion segmentation is called **change detection** (briefly in the next slides)
  - **tracking:** estimate the (2D or even 3D) trajectories of points or objects from image sequences (next week)

CHANGE DETECTION
(IN A FEW WORDS)

I_t

M_t



- in the case the camera is still motion segmentation can be implemented as a difference operation
- assuming we have a reference image

$$M_t(x,y) = \begin{cases} 1 & \text{if } |I_t(x,y) - I_{ref}(x,y)| > \tau \\ 0 & \text{otherwise} \end{cases}$$

CHANGE DETECTION

WHAT IS A REFERENCE IMAGE?

- the reference image or, more in general the *background model,* is a "picture" of the empty scene, containing all the parts which are not moving

- the simplest way of computing it is to average the first N frames (assuming that at the beginning the scene is stationary)

$$I_{ref} = B = \frac{1}{N}\sum_{t=1}^{N} I_t$$

MODELING A BACKGROUND BY A **RUNNING AVERAGE**

$$B_t(x,y) = \begin{cases} B_{t-1}(x,y) & \text{if } |I_t(x,y) - B_{t-1}(x,y)| > \tau \\ (1-\alpha)B_{t-1}(x,y) + \alpha I_t(x,y) & \text{otherwise} \end{cases}$$

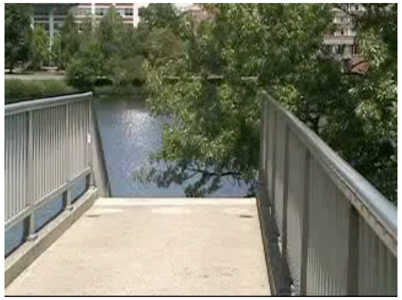
notice: the background model is updated at each frame

PROS

- it is quite robust to moving objects in the scene
- it incorporates stable changes (at a speed which is proportional to α)
- it is simple and computationally efficient

CONS

- it does not deal with repetitive (and uninteresting) motion



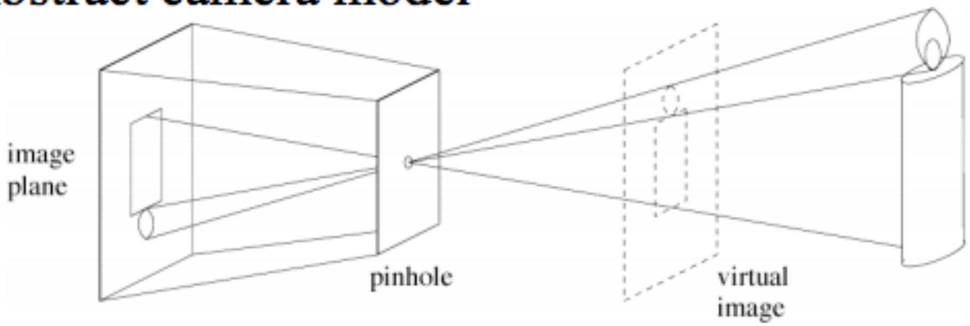
CORRESPONDENCE PROBLEM

- if the camera is still, correspondence is trivial, but in general…
- **correspondence:** which elements of a frame correspond to which elements of the next frame?
- in the case of motion, correspondence may also be seen as a problem of *estimating the apparent motion of the brightness pattern* (the so called **optical flow**)

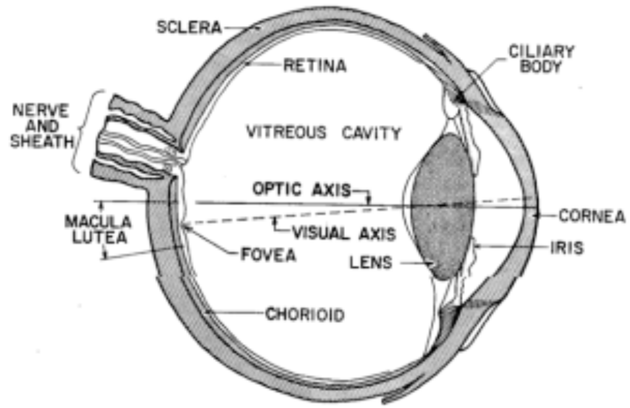
REMINDER

PINHOLE CAMERA

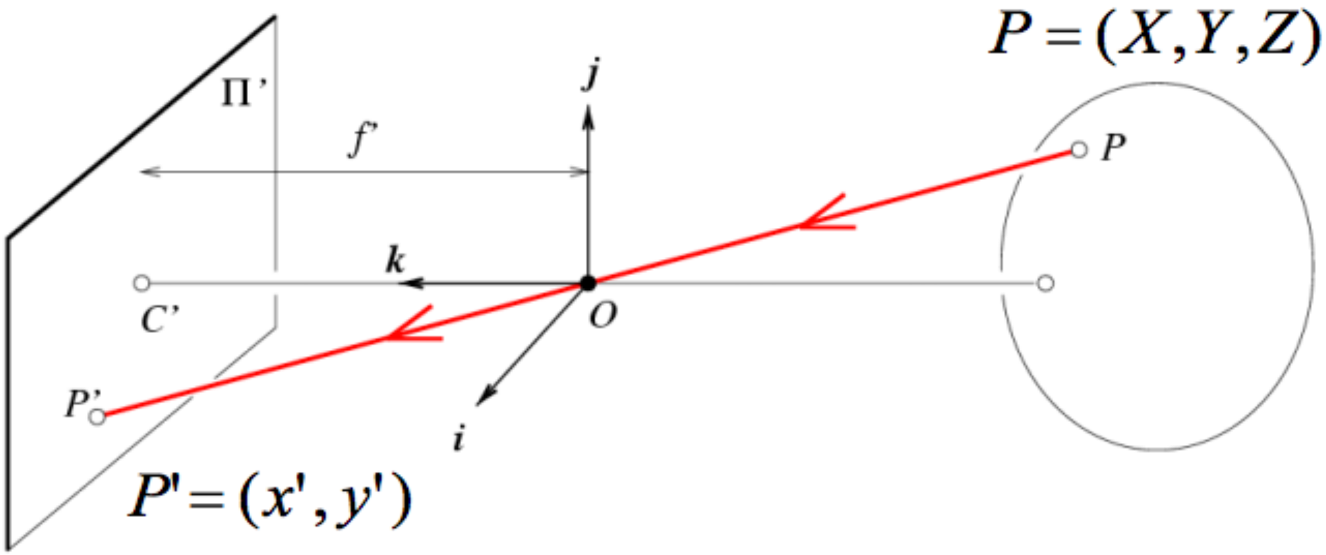
Abstract camera model



Animal eye

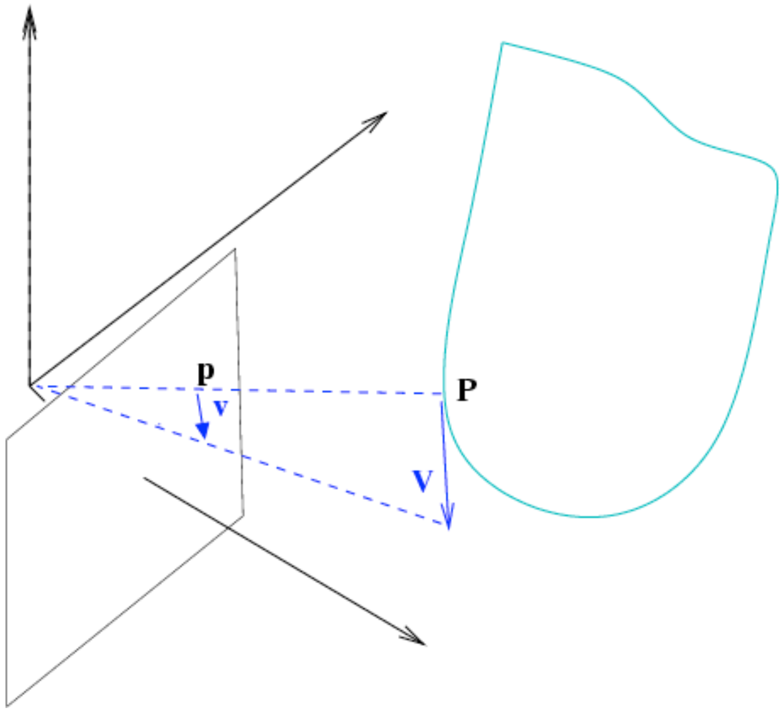


Pinhole Perspective Equation



$$\begin{cases} x' = f' \dfrac{X}{Z} \\ y' = f' \dfrac{Y}{Z} \end{cases}$$

THE MOTION FIELD



- the motion field is the 2D vector field of velocities of the image points, induced by the relative motion between camera and scene

- it can be seen as the projection of the 3D velocity field on the image plane

THE MOTION FIELD OF RIGID OBJECTS

- rigid objects have a **constant** 3D velocity field
- let us consider a 3D point $\mathbf{X}=(X,Y,Z)^T$
- the relative motion between $\mathbf{X}$ and the camera will have a translational and an angular component

$$\mathbf{V} = \underbrace{-\mathbf{W}}_{\text{translational}} - \underbrace{\mathbf{\Omega}}_{\text{angular}} \times \mathbf{X}$$

motion of the object
w.r.t the camera
(hence the negative signs)

THE MOTION FIELD

- the motion field **v** is the projection of **V** on the image plane
- the motion field **v** is the velocity (time derivative) of point **x** projection of **X**

$$\mathbf{v} = \dot{\mathbf{x}} = \frac{\partial \mathbf{x}}{\partial t} = f \frac{\partial}{\partial t}(\frac{\mathbf{X}}{Z})$$

$$\mathbf{v} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ 0 \end{pmatrix} = f \begin{pmatrix} \frac{Z\dot{X} - \dot{Z}X}{Z^2} \\ \frac{Z\dot{Y} - \dot{Z}Y}{Z^2} \\ 0 \end{pmatrix} = f \frac{Z\dot{\mathbf{X}} - \dot{Z}\mathbf{X}}{Z^2} = f \frac{Z\mathbf{V} - V_Z\mathbf{X}}{Z^2}$$

EQUATIONS OF THE MOTION FIELD

- Summing up....

$$
\begin{aligned}
\mathbf{V} &= -\mathbf{W} - \boldsymbol{\Omega} \times \mathbf{X} \\
\mathbf{v} &= f\frac{Z\mathbf{V} - V_Z\mathbf{X}}{Z^2}
\end{aligned}
$$

- we obtain

$$
\mathbf{v} = \underbrace{-f\frac{\mathbf{W}}{Z} + \frac{W_Z\mathbf{x}}{Z}}_{\text{translational component}} \underbrace{- \boldsymbol{\Omega} \times \mathbf{x} + (\boldsymbol{\Omega} \times \frac{\mathbf{x}}{f})^\top \hat{k}\mathbf{x}}_{\text{rotational component}}
$$

TWO IMPORTANT OBSERVATIONS

- Angular velocity and 3D point depth never appear together
  - this means that with pure rotation the motion field does not depend on the scene structure
- Translational velocity $\mathbf{W}$ is always divided by Z: *depth-velocity ambiguity*
  - this means that the same motion field can be obtained by a close object moving slowly or a far object moving fast

SPECIAL MOTION: PURE TRANSLATION

$$\begin{pmatrix} v_x \\ v_y \\ 0 \end{pmatrix} = \begin{pmatrix} -f\frac{W_X}{Z} + \frac{W_Z x}{Z} \\ -f\frac{W_Y}{Z} + \frac{W_Z y}{Z} \\ 0 \end{pmatrix}$$

- we observe there exists a point p_0 whose velocity is instantaneously 0

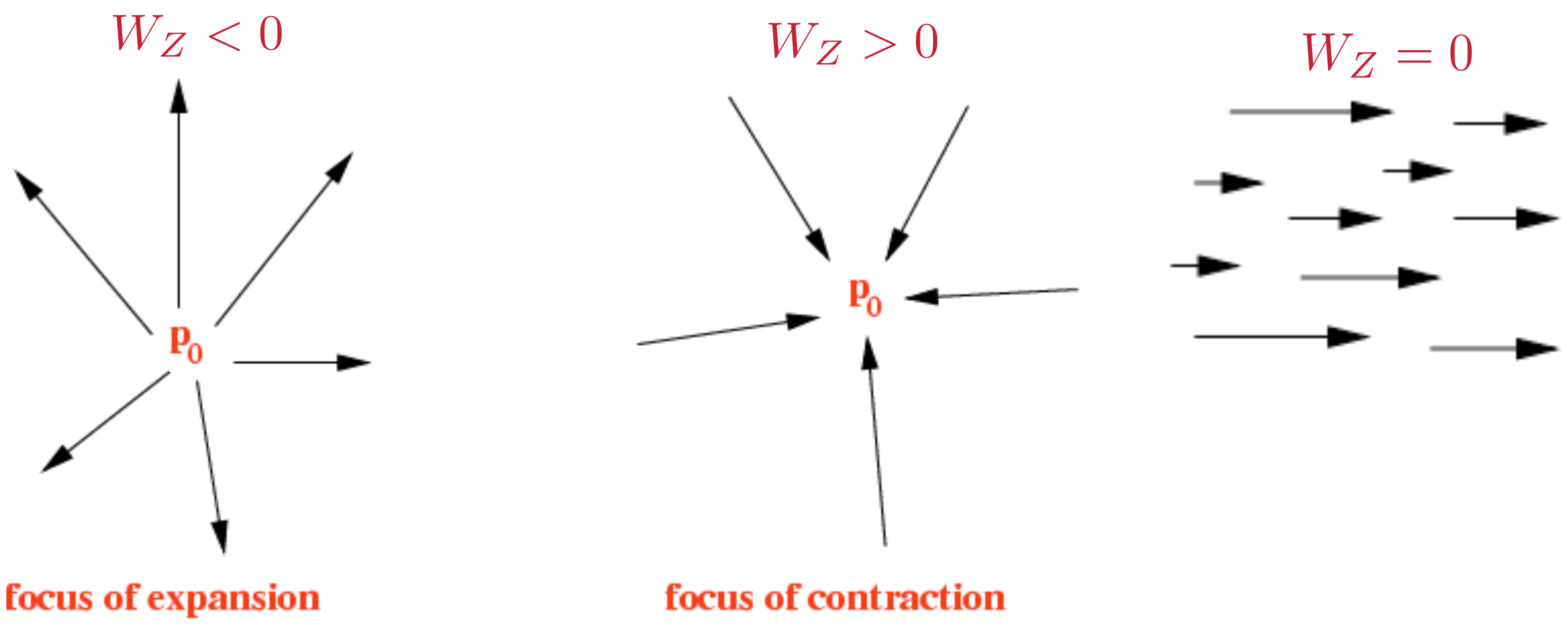
$$\begin{aligned} x_0 &= f\frac{W_X}{W_Z} \\ y_0 &= f\frac{W_Y}{W_Z} \end{aligned}$$

- thus we may write

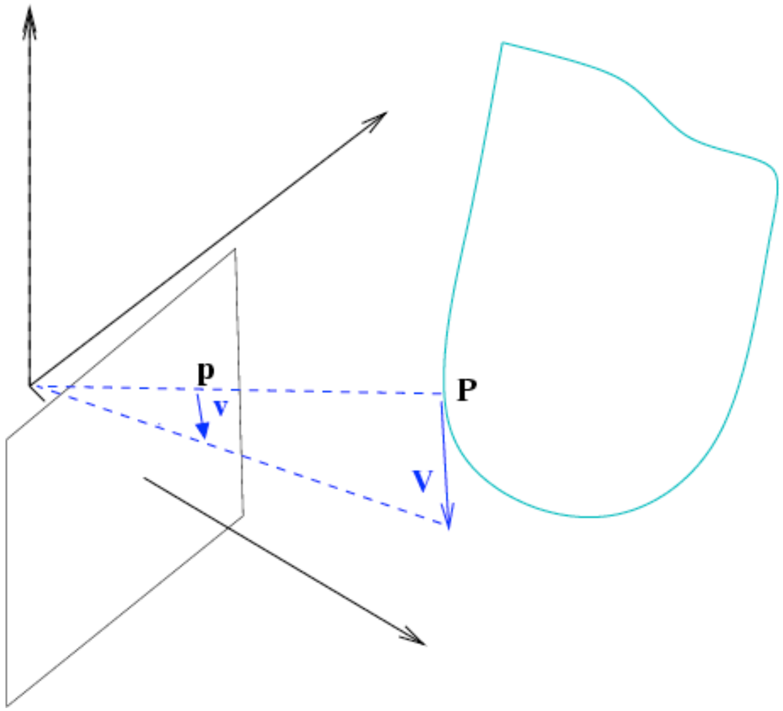
$$\begin{pmatrix} v_x \\ v_y \\ 0 \end{pmatrix} = \begin{pmatrix} (x - x_0)\frac{W_Z}{Z} \\ (y - y_0)\frac{W_Z}{Z} \\ 0 \end{pmatrix}$$

SPECIAL MOTION: PURE TRANSLATION

- in the case of pure rotation the motion field converges to a point p_0 called the focus of expansion



THE MOTION FIELD



- the motion field is the 2D vector field of velocities of the image points, induced by the relative motion between camera and scene

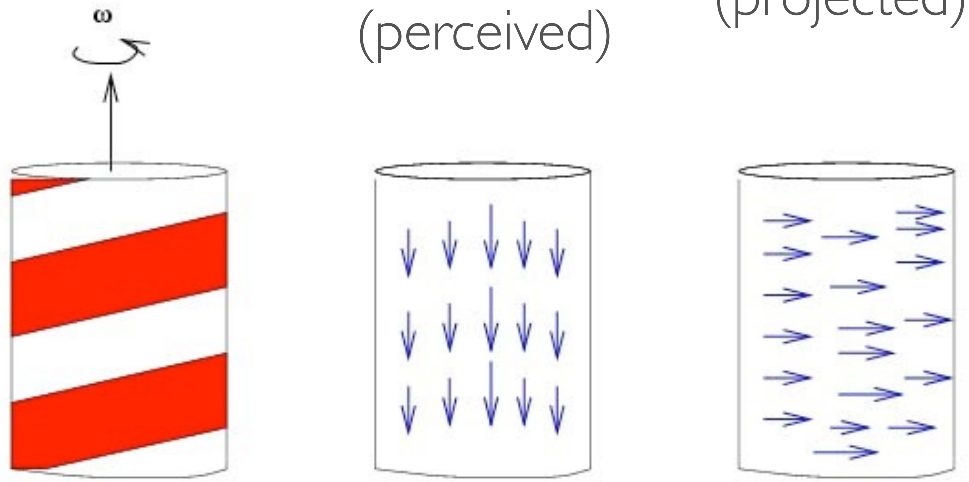
- it can be seen as the projection of the 3D velocity field on the image plane

MOTION FIELD AND ITS ESTIMATE

- we may estimate the motion field from images, but what we estimate will be related to the *apparent* motion

- what is the motion field of an object moving in a dark room? or of a uniform object on a similar background?

3D motion (rotational) | apparent motion (perceived) | motion field (projected)



HOW TO ESTIMATE THE MOTION FIELD

- we need to make a strong assumption on the *image brightness constancy*

$$\frac{dI}{dt} = 0$$

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

$$(\nabla I)^\top \mathbf{u} + I_t = 0$$

image brightness constancy equation

OPTICAL FLOW

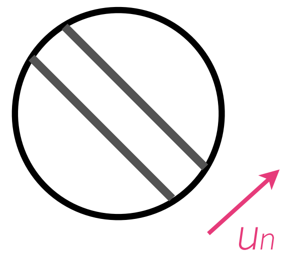
- The optical flow is a vector field subject to the constraint

$$(\nabla I)^\top \mathbf{u} + I_t = 0$$

- Notice that **one** constraint is not enough to compute the optical flow (**2** unknowns)

THE APERTURE PROBLEM

- the image brightness constancy equation allows us to determine the optical flow component parallel to the spatial image gradient

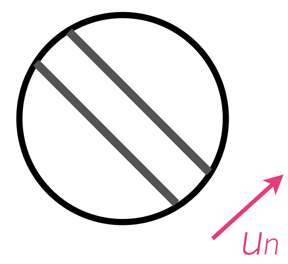


- Analytically

$$u_n = \frac{(\nabla I)^\top \mathbf{u}}{||\nabla I||} = \frac{-I_t}{||\nabla I||}$$

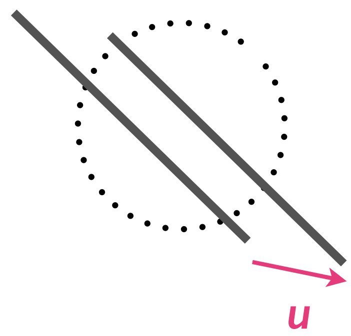
THE APERTURE PROBLEM

- the image brightness constancy equation allows us to determine the optical flow component parallel to the spatial image gradient



- Analytically

$$u_n = \frac{(\nabla I)^\top \mathbf{u}}{||\nabla I||} = \frac{-I_t}{||\nabla I||}$$



OPTICAL FLOW ALGORITHMS

- many algorithms start from the idea of adding constraints to the underdetermined system obtained by the brightness constancy equation
- we will see a simple way of doing so: the Lucas-Kanade algorithm
    - assumption: **u** is <u>constant</u> in a small neighbourhood of a point

LUCAS KANADE ALGORITHM

- the assumption allows us to obtain a system of equations with one equation for each point in the neighbourhood

$$(\nabla I(\mathbf{x}_i, t))^\top \mathbf{u} + I_t(\mathbf{x}_i, t) = 0 \quad \mathbf{x}_i \in N$$

- we then obtain a linear system **Au=b** with

$$A = \begin{bmatrix} \nabla I(\mathbf{x}_1, t)^\top \\ \nabla I(\mathbf{x}_2, t)^\top \\ \cdot \\ \cdot \\ \cdot \\ \nabla I(\mathbf{x}_m, t)^\top \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} -I_t(\mathbf{x}_1, t) \\ -I_t(\mathbf{x}_2, t) \\ \cdot \\ \cdot \\ \cdot \\ -I_t(\mathbf{x}_m, t) \end{bmatrix}$$

m elements in the *N* neighbour.

LUCAS KANADE ALGORITHM

- The linear system may be solved with the pseudo-inverse

$$\mathbf{u} = A^{\dagger}\mathbf{b} \quad \text{with} \quad A^{\dagger} = (A^{\top}A)^{-1}A^{\top}$$

-

LUCAS KANADE ALGORITHM

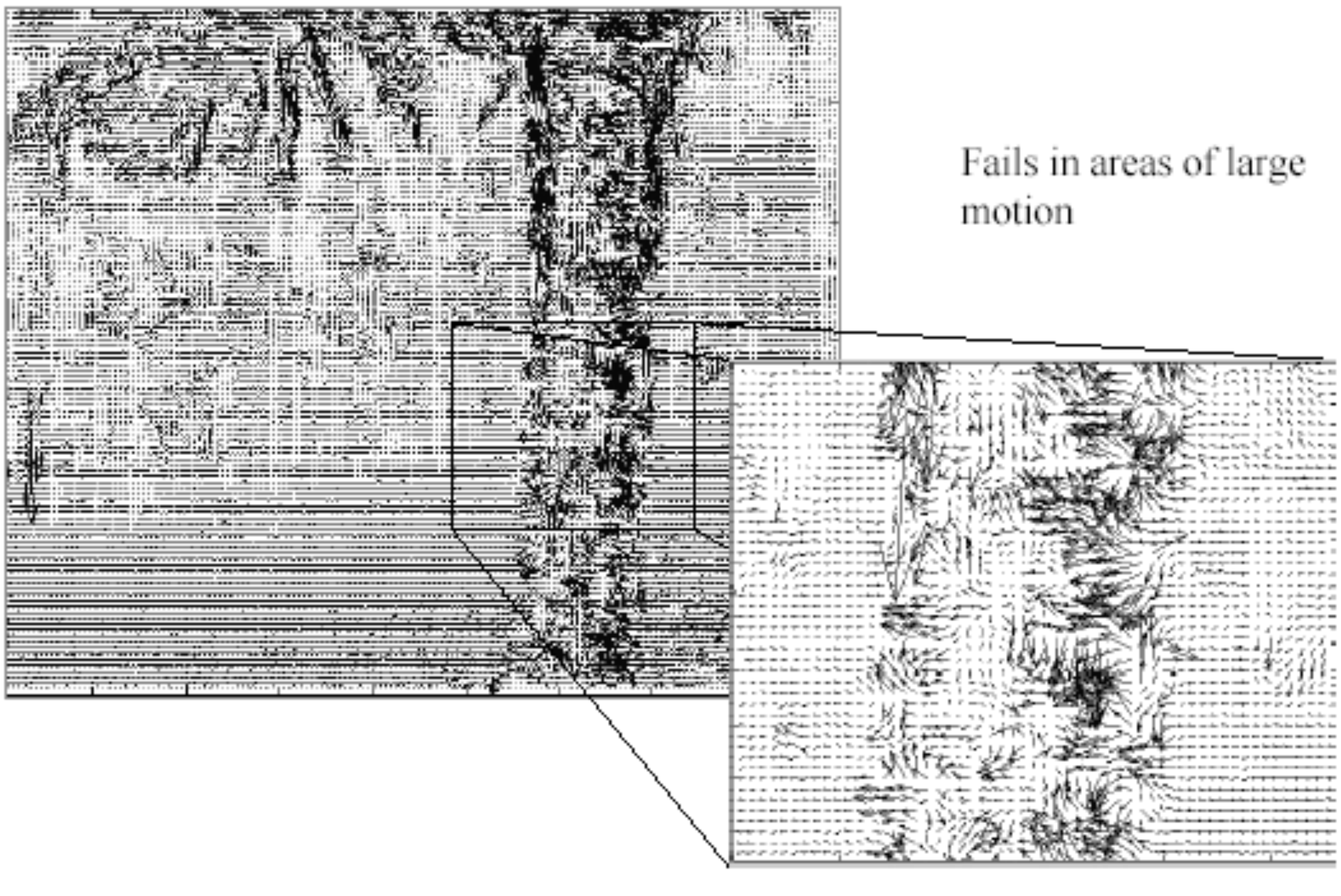
- The linear system may be solved with the pseudo-inverse

$$\mathbf{u} = A^{\dagger}\mathbf{b} \quad \text{with} \quad A^{\dagger} = (A^{\top}A)^{-1}A^{\top}$$

- notice that the inversion of the matrix will be ill-posed if the matrix is not full rank (corners)

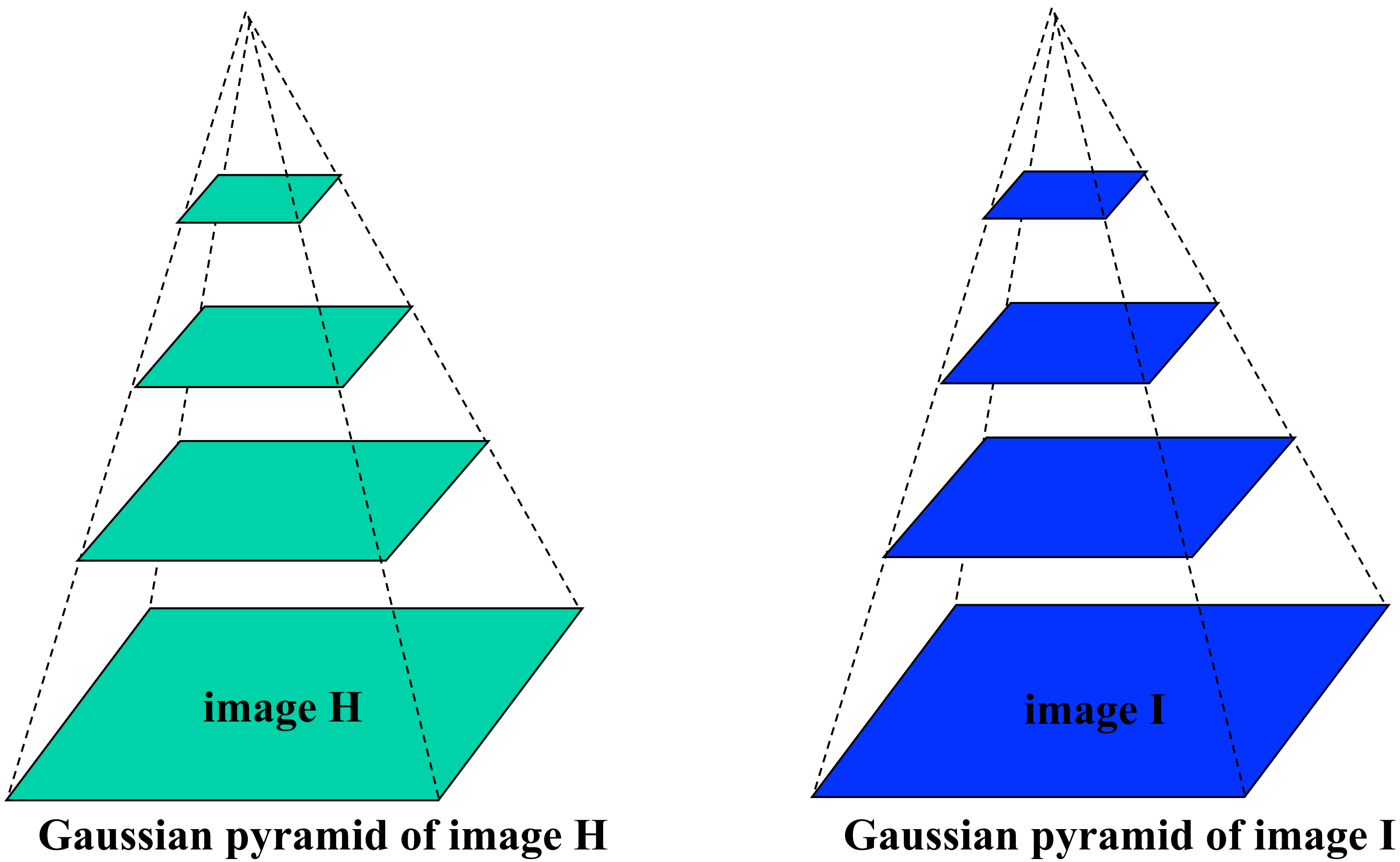
$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

HOW GOOD ARE THE RESULTS?

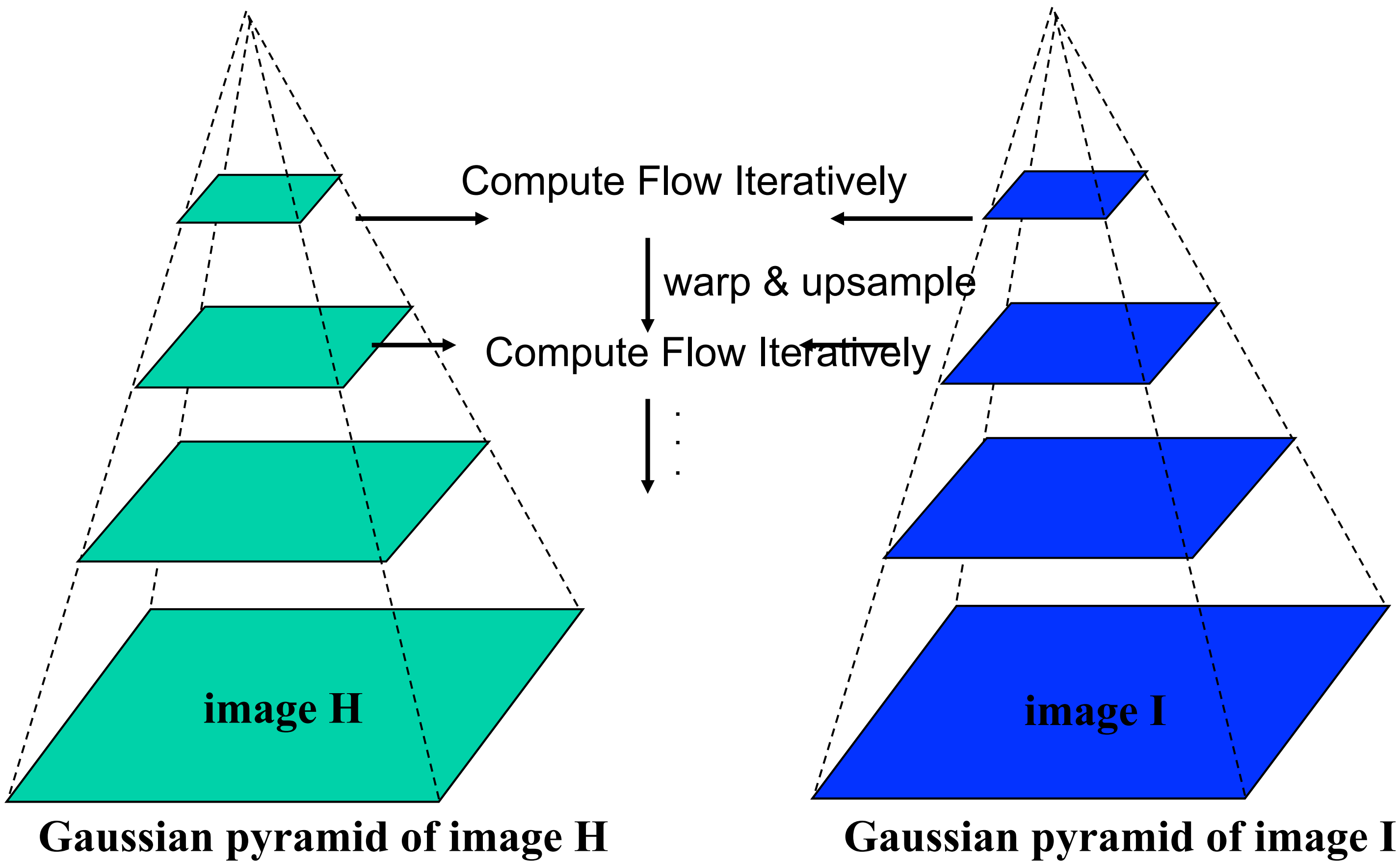


Fails in areas of large motion

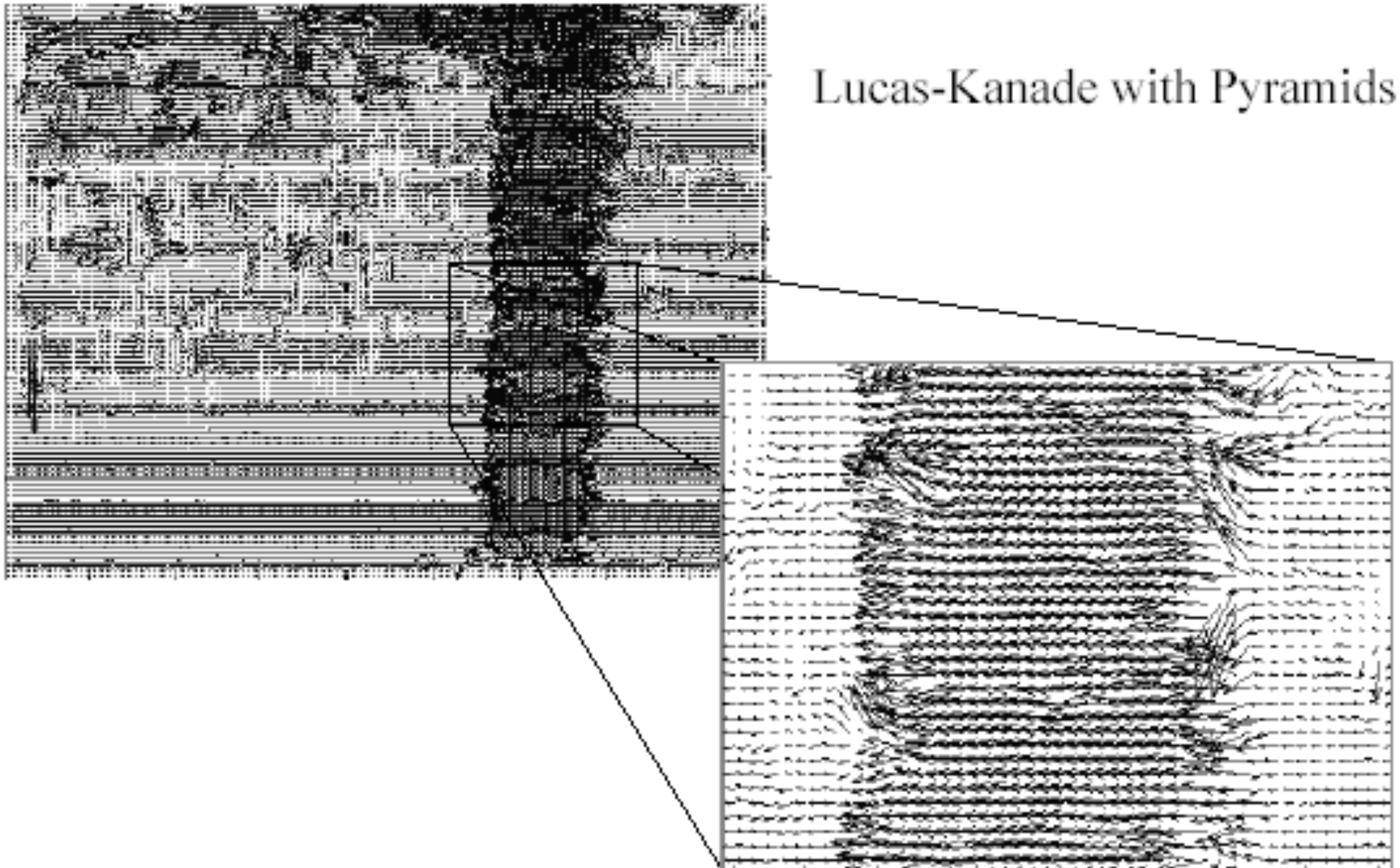
COARSE TO FINE ESTIMATION



COARSE TO FINE ESTIMATION



COARSE TO FINE ESTIMATION: RESULTS



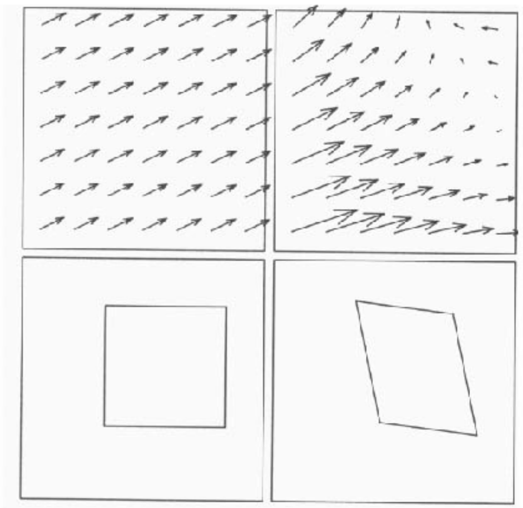
Lucas-Kanade with Pyramids

OTHER PARAMETRIC MODELS

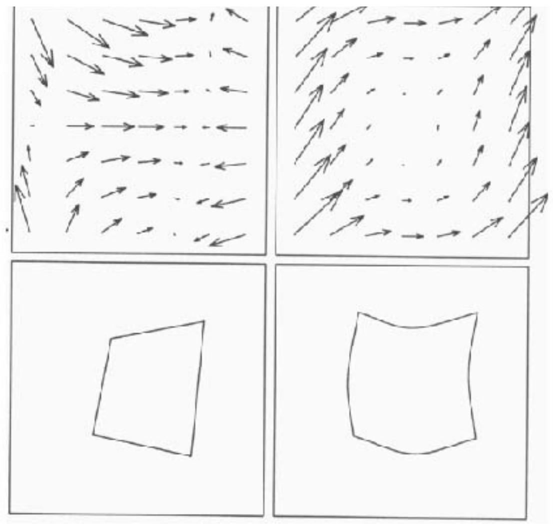
- the hypothesis of a locally constant (translational) optical flow can be extended with more complex models (affine, projective, quadratic, ...)

the number of unknowns increase!

translational (constant) | affine

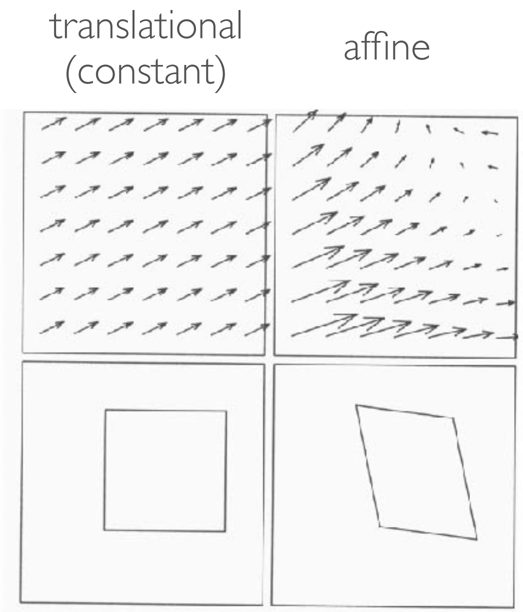


perspective | quadratic



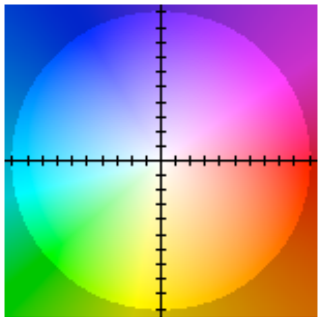
OTHER PARAMETRIC MODELS

- the hypothesis of a locally constant (translational) optical flow can be extended with more complex models (affine, projective, quadratic, ...)



affine assumption:
all points q in a neighbourhood of p obey to
$u_q = M u_p + n$

EXAMPLES - VISUALIZATION



EXAMPLES - VISUALIZATION

APPLICATIONS TO RELATED PROBLEMS

object tracking