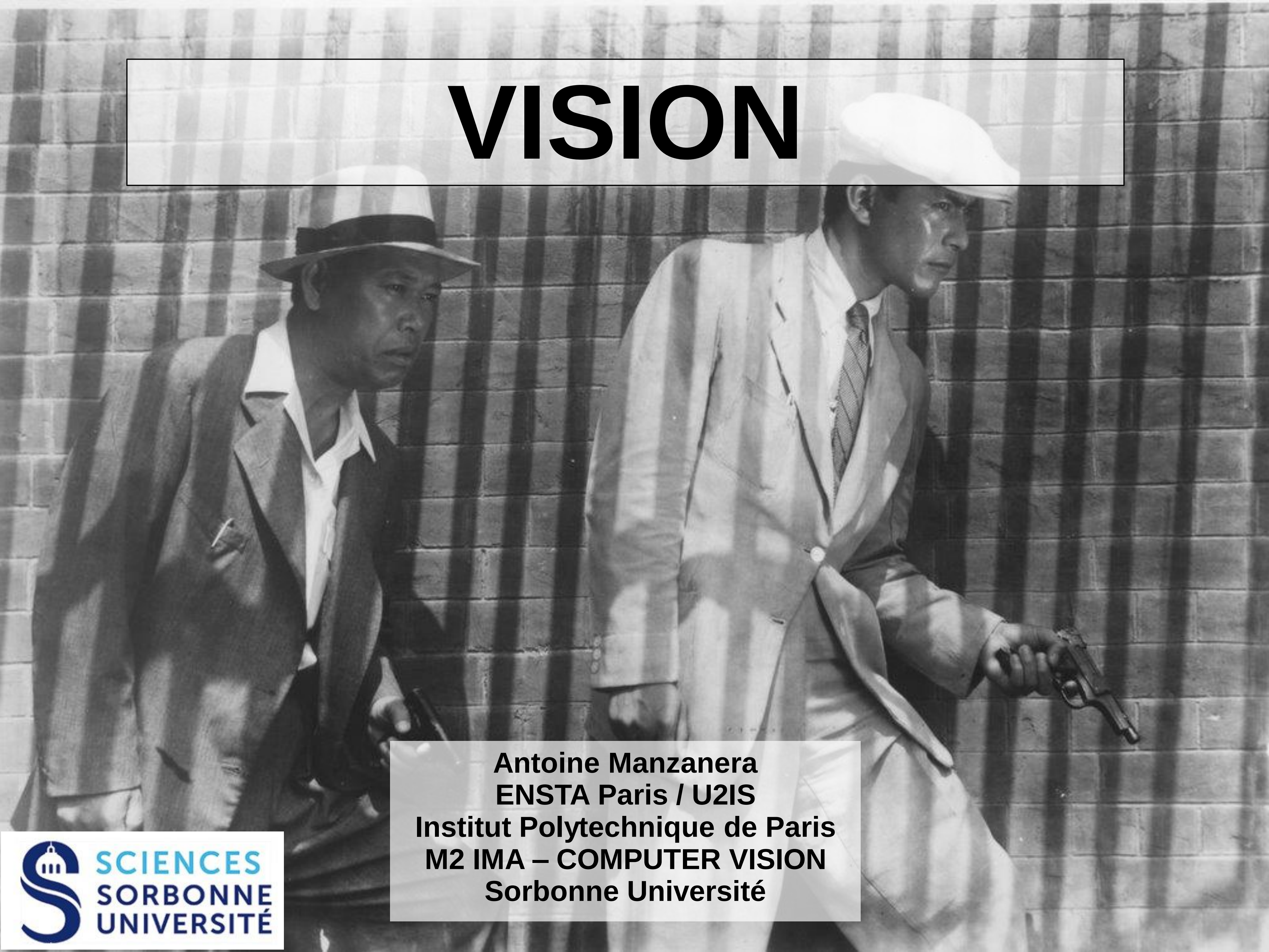
VISION



Antoine Manzanera
ENSTA Paris / U2IS
Institut Polytechnique de Paris
M2 IMA – COMPUTER VISION
Sorbonne Université

SCIENCES SORBONNE UNIVERSITÉ

VISION COURSE #2:
Some co-design techniques for motion analysis and 3d reconstruction in Computer Vision

Objectives of the lecture:

- Getting a global view of the co-design opportunistic approaches, making the most of the different parts of a computer vision system (optics / mechanics / electronics / software) to increase its perception and analysis capabilities.
- Understanding the principle of the main categories of co-design approaches in Computer vision.

SCIENCES SORBONNE UNIVERSITÉ

VISION COURSE #2:
Some co-design techniques for motion analysis and 3d reconstruction in Computer Vision

OUTLINE OF THIS LECTURE:

- **Part 1, Active 3d:**
  - Time-of-Flight Cameras
  - Structured Light
- **Part 2, Passive 3d:**
  - Plenoptic cameras
  - Depth from focus
  - Coded aperture
- **Part 3, Electronic retinas:**
  - Event-based cameras
  - Programmable retinas

Part 1: 3D CAMERAS / ACTIVE APPROACHES

Active 3d cameras aim at measuring the depth of every point from the scene that is projected on the image plane, using its response to a particular lighting.

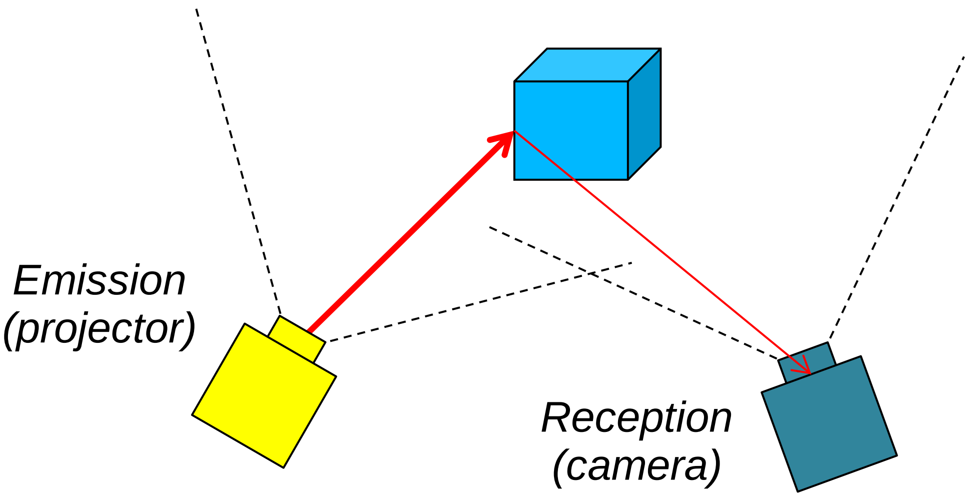
The two fundamental components of the system are then:

1. A lighting system controlled in time and space
2. A sensing device to analyse the illuminated scene

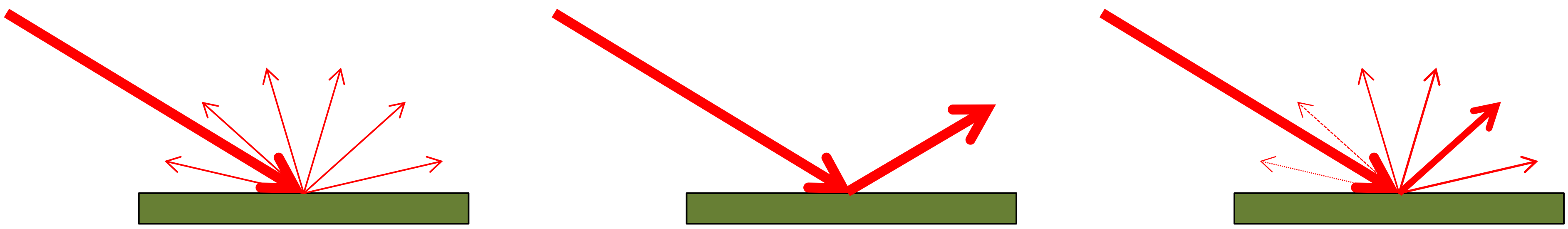
Such systems are active in that they *emit* a light signal (not to be confused with the other sense of « active vision », i.e. that « moves to see »).

ACTIVE APPROACHES AND DIFFUSION MODELS

For 3d active cameras, it is assumed that every point illuminated by the projector in the camera field of view reflects a part of its light toward the optical centre of the camera.



The nature of light diffusion at the measured point then has a major influence in the depth estimation…



Lambertian diffusion

Perfect specular reflection (mirror)

Semi specular diffusion

Note that this may also be an issue for passive approaches (e.g point matching between two poses).

ACTIVE 3D: “TIME OF FLIGHT” CAMERAS

3d « time of flight » (ToF) cameras measure the distance d_x between a point X projected in x, from the propagation time t_x of light (with speed c) from its emission by the projector until its reception by the photosensor associated to x, after being reflected by point X:

$$d_x = \frac{c.t_x}{2}$$

Unlike scanner like (e.g. LIDAR) systems, the light emitted by ToF cameras (usually laser infrared LED) illuminates the whole scene simultaneously.

Different technologies can be used to measure time of flight:

- Direct time measuring (impulsion light)
- Phase estimating (time-modulated continuous light).

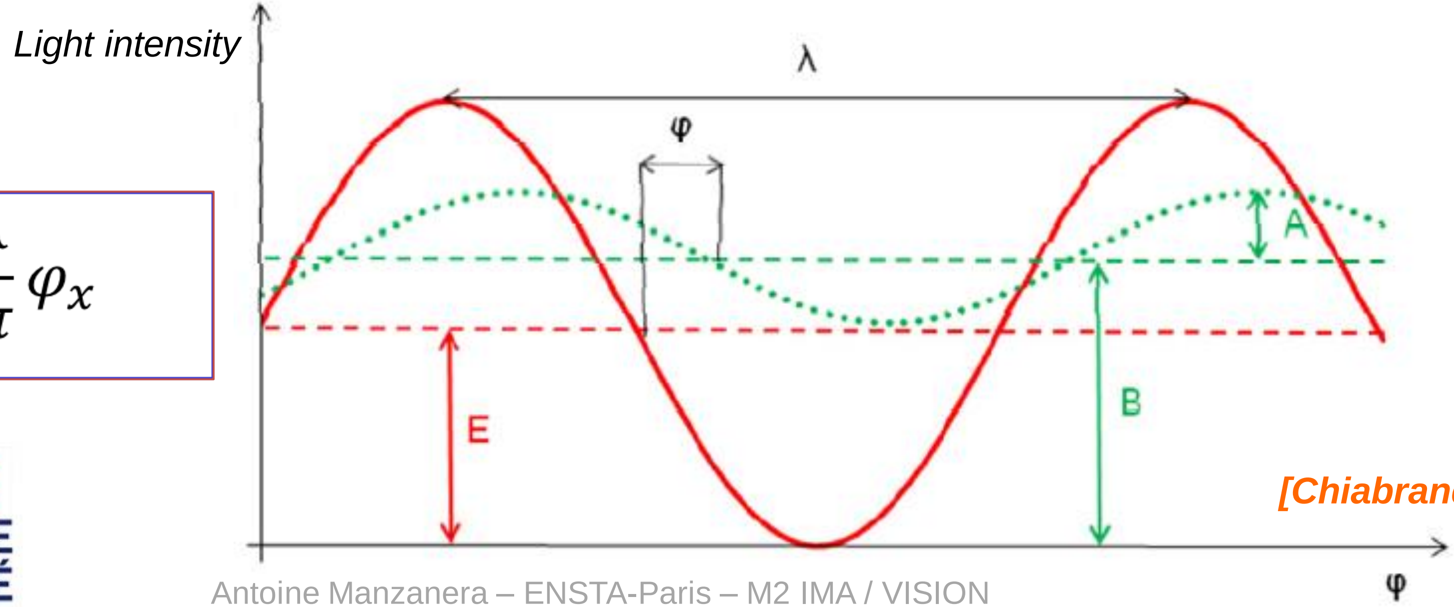
[CamCube - ©PMDTech]

[Kinect v2 for Xbox One - ©Microsoft]

ACTIVE 3D: ToF CAMERA BASED ON PHASE ESTIMATION

- The scene is uniformly illuminated with a light whose intensity varies in time according to a sine signal (in red) with amplitude E.
- The signal received in pixel *x* (in green) has the same frequence, a weaker amplitude A depending on the reflectivity of the point and a phase shift φ depending on its distance.
- The signal received is also shifted in intensity (offset) of a value B due to the background light present in the scene.
- This signal is sampled and the phase shift φ is deduced from the measured intensities.
- The modulation period λ (typ. 50 ns) is large with respect to the time of flight to avoid phase ambiguities, but small with respect to typical acquisition times to allow repeating the measure (time filtering).

$$d_x = \frac{c\lambda}{4\pi}\varphi_x$$



[Chiabrando 2009]

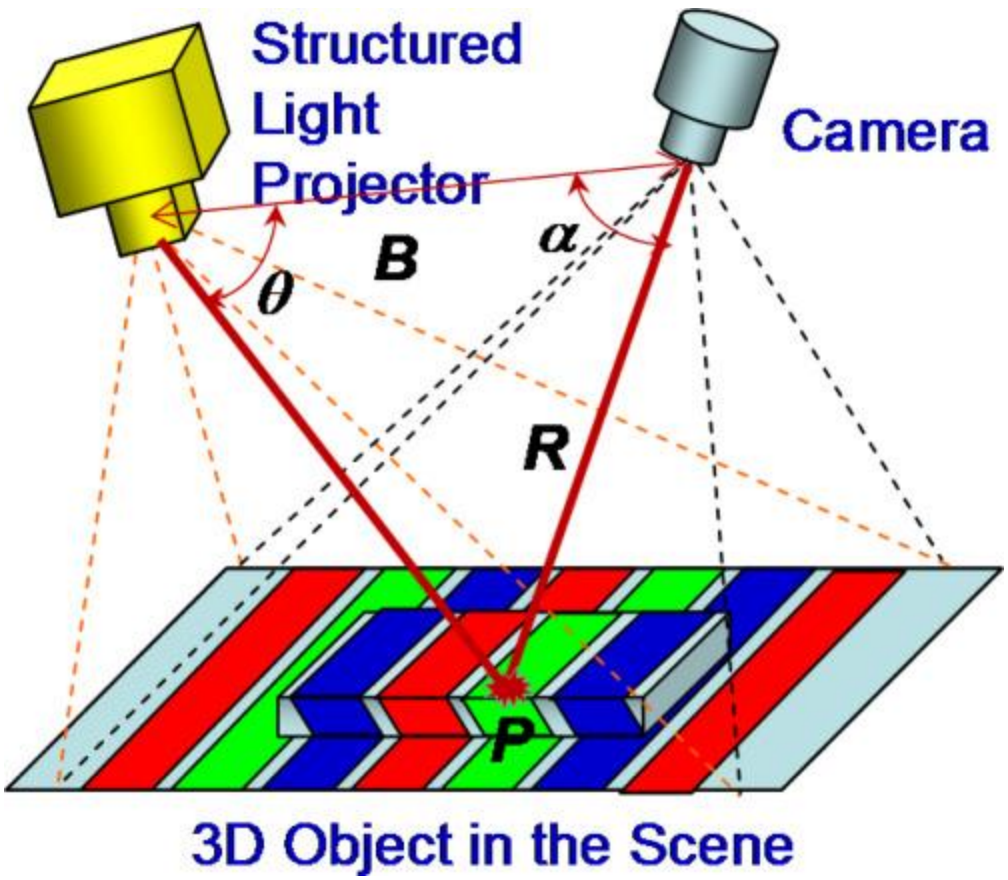
ACTIVE 3D: "STRUCTURED LIGHT" CAMERAS

Structured light 3d cameras interpret the deformation of a 2d image projected into the scene to recover depth information.

They are based on the same triangulation principle as stereovision:

$$R = B \frac{\sin\theta}{\sin(\alpha + \theta)}$$

The structure of projected 2d images determines a spatial coding that plays a major role in triangulation.

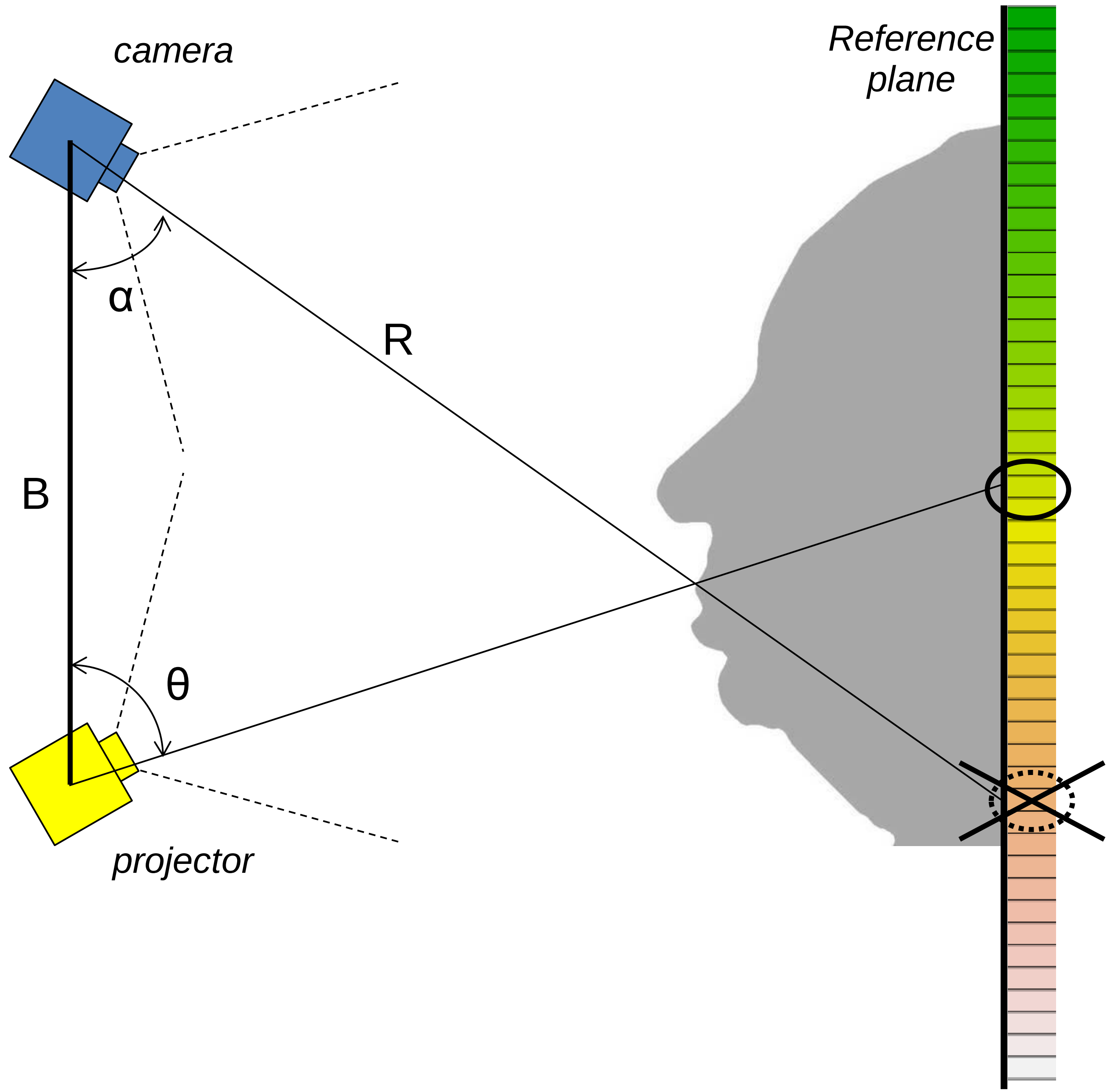


[Geng 2011]

ACTIVE 3D: "STRUCTURED LIGHT" CAMERAS

$$R = B\frac{\sin\theta}{\sin(\alpha+\theta)}$$

The angle α is provided by the position of the point in the image, and the angle θ by the corresponding colour (or pattern) in the reference plane:



ACTIVE 3D: "STRUCTURED LIGHT" CAMERAS

Also:

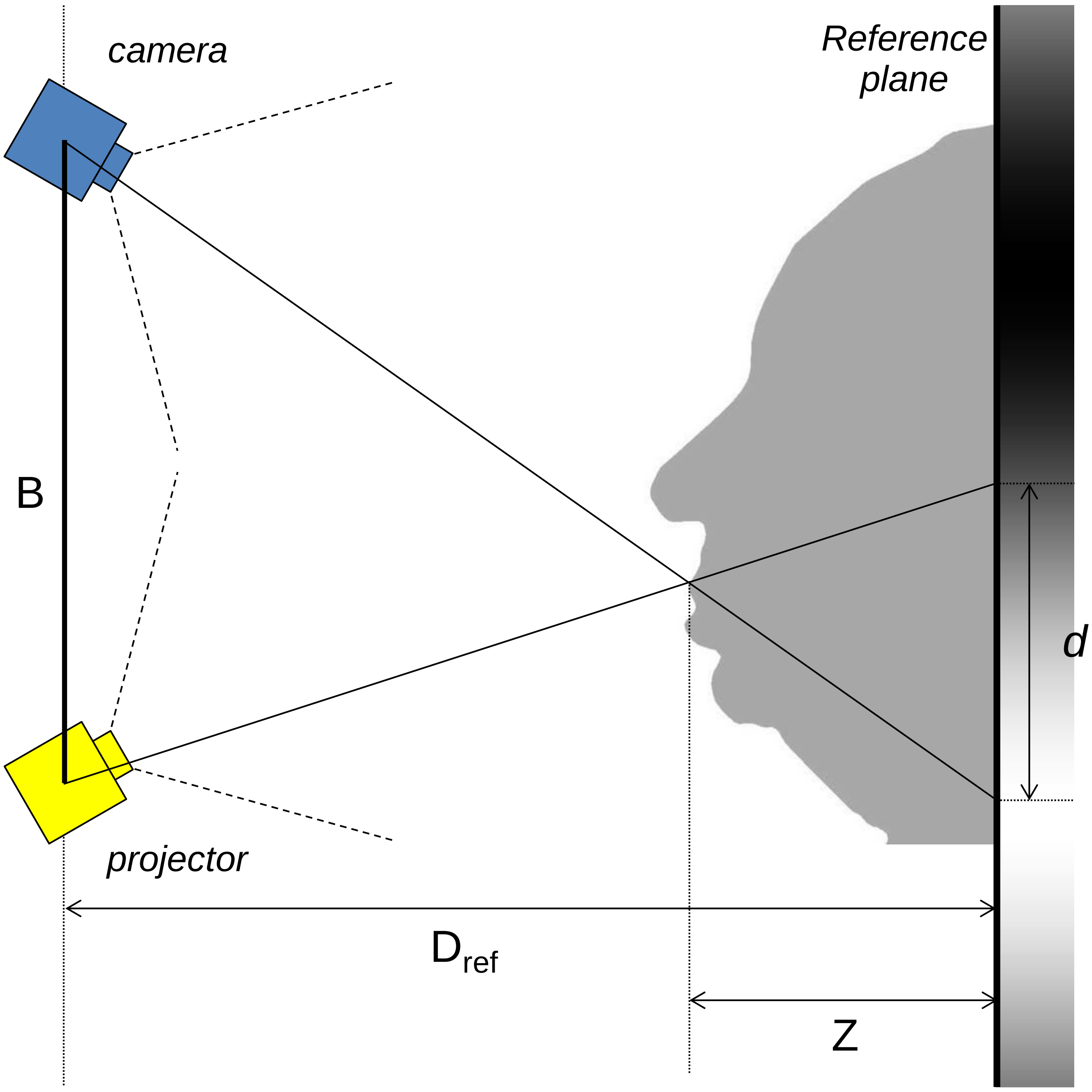
$$\frac{d}{B} = \frac{Z}{D_{ref} - Z}$$

And then:

$$Z \approx \frac{D_{ref}}{B} d$$

So, if the projected image is a sinusoidal ramp, depth can be deduced from the phase shift:

$$Z \propto \Delta\varphi$$

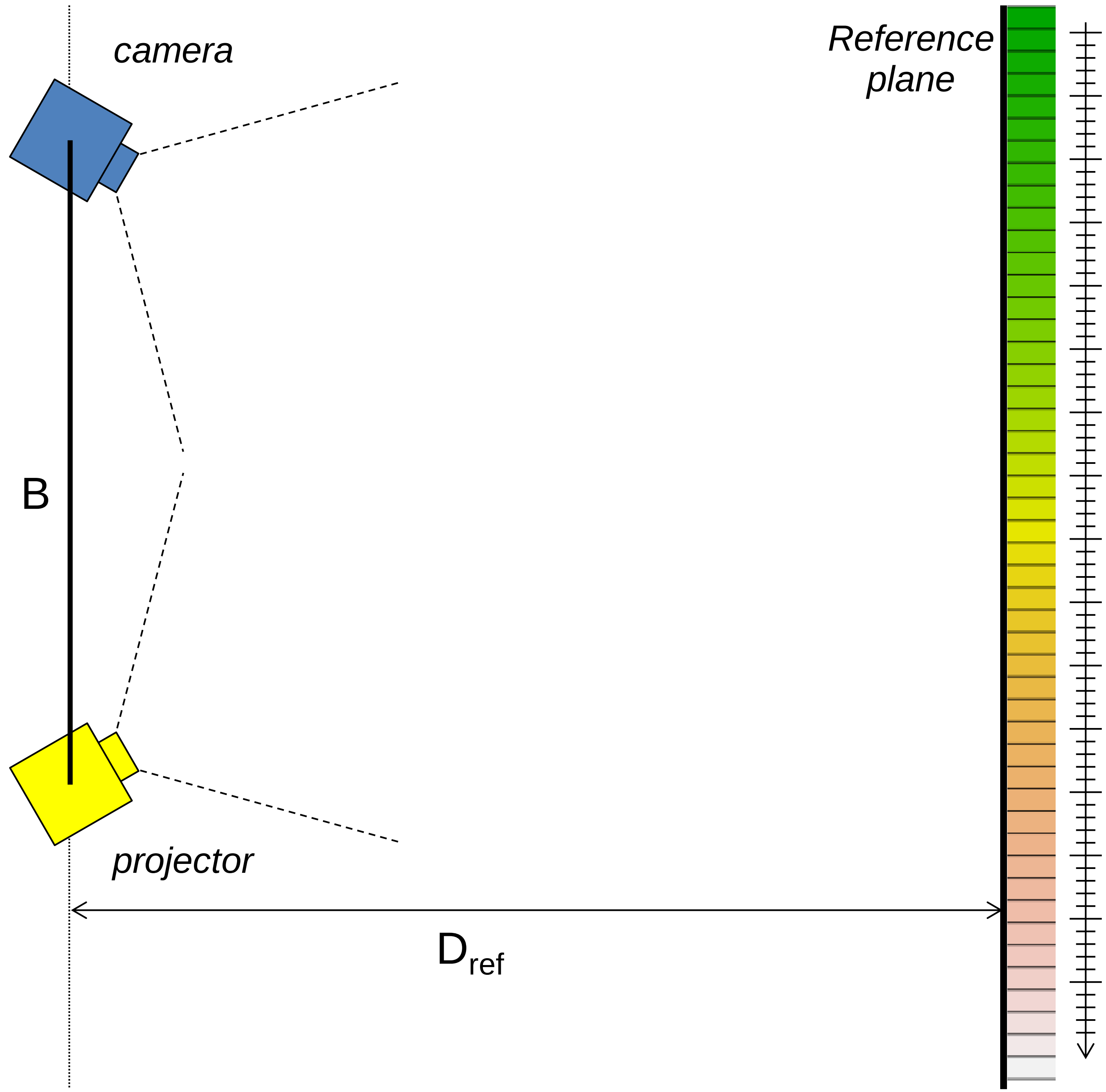


"STRUCTURED LIGHT" CAMERAS: CALIBRATION

Like stereo systems, the camera and the projector must be calibrated to:

(1) Determine the back-projection line for every pixel of the captured image.

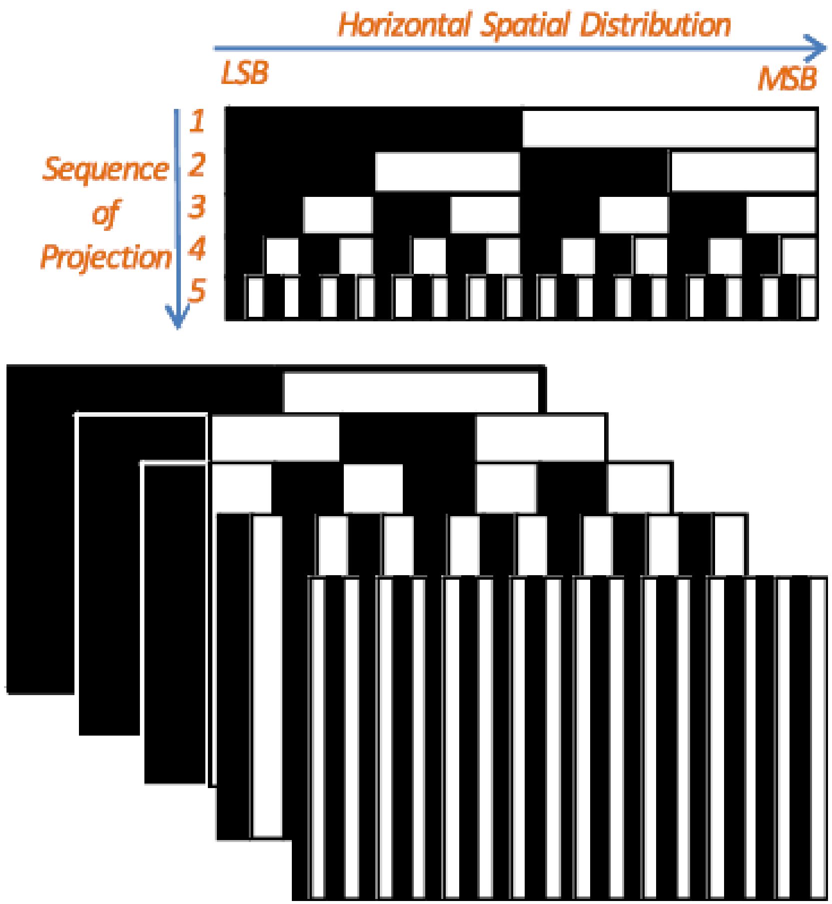
(2) Associate to each pattern of the projected image the direction corresponding to its projection on the reference plane.



STRUCTURED LIGHT: WHICH PATTERNS?

- Ideally, every point should be uniquely indentified from its value/colour…
  - …but all the values must be easily distinguishable!
- A point can also be identified using its neighbourhood…
  - …but then each neighbourhood must be unique!
- Depth being associated to an angle, a 1d target (band) is sufficient…
  - ..but using a 2d target may solve ambiguities!
- Several targets may also be sequentially combined...
  - …but then the acquisition time increases!

STRUCTURED LIGHT: SEQUENTIAL TARGETS



Binary sequence 2^5

[Posdamer 1982, from Geng 2011]

Binary targets allow to optimally discriminate the different values.

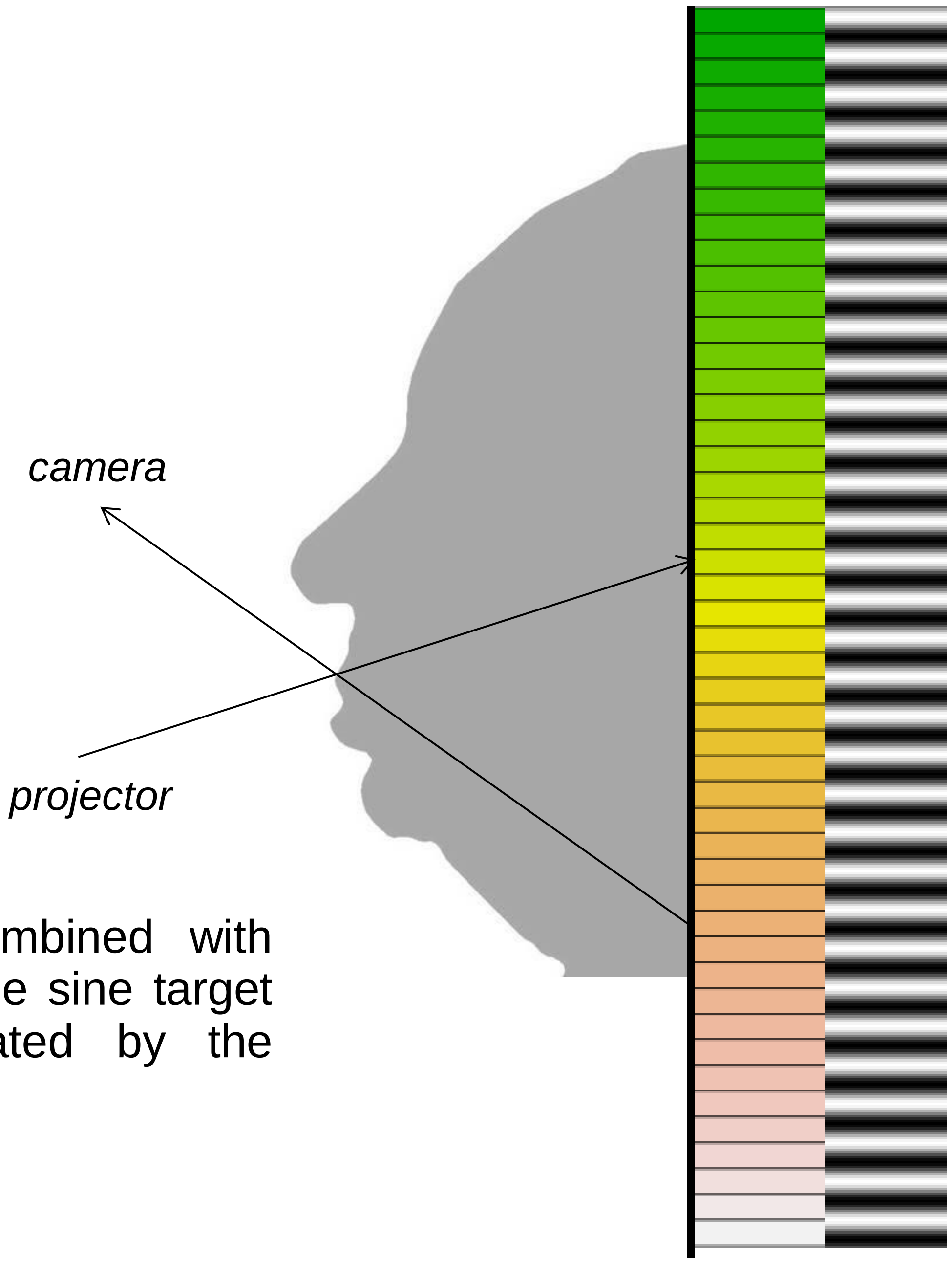
Depth resolution depends on the number of distinct values and then, for sequential techniques, on the acquisition time.

[from Naramsimhan 2006]

STRUCTURED LIGHT: SEQUENTIAL TARGETS

Increasing the number of bits for a trade-off contrast / acquisition time:

Ternary sequence 3^3

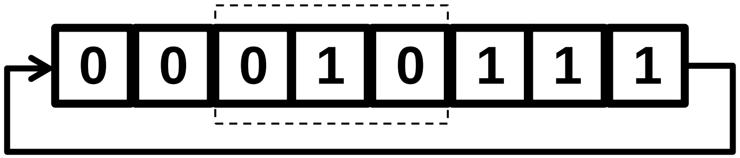


Rectangular signal targets can also be combined with sinusoid targets (on the right): the phase of the sine target allows to refine the coarse depth estimated by the rectangular target.

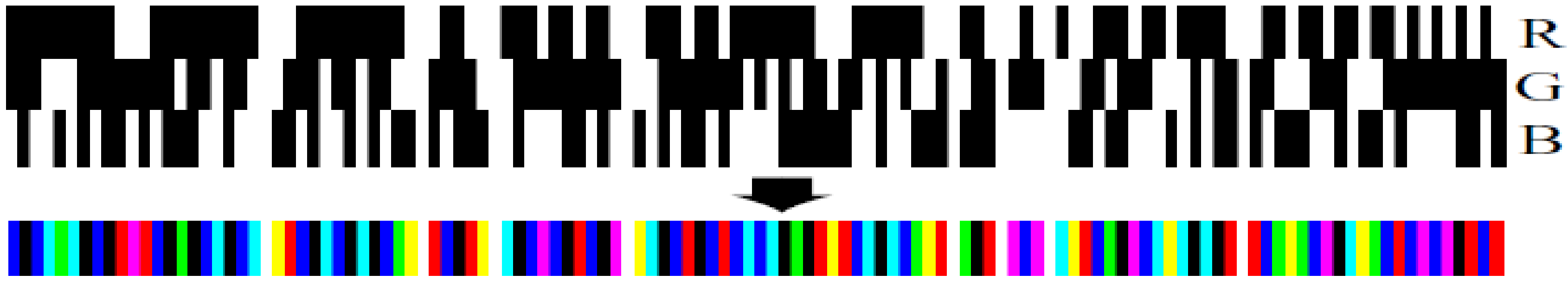
STRUCTURED LIGHT: UNIQUE “SNAPSHOT” TARGET

- To better distinguish the values, rectangular (runs) targets are preferred to continuous ones (ramps).
- To be able to locally discriminate points using quantised values, local patterns (neighbourhoods) can be used instead of the value alone.
- But then, each pattern must define a *unique* position.

De Bruijn’s sequence *B(n,k)* are words from a *n*-symbols alphabet such that all the sub-words of length *k* are different.



De Bruijn’s sequence B(2,3)

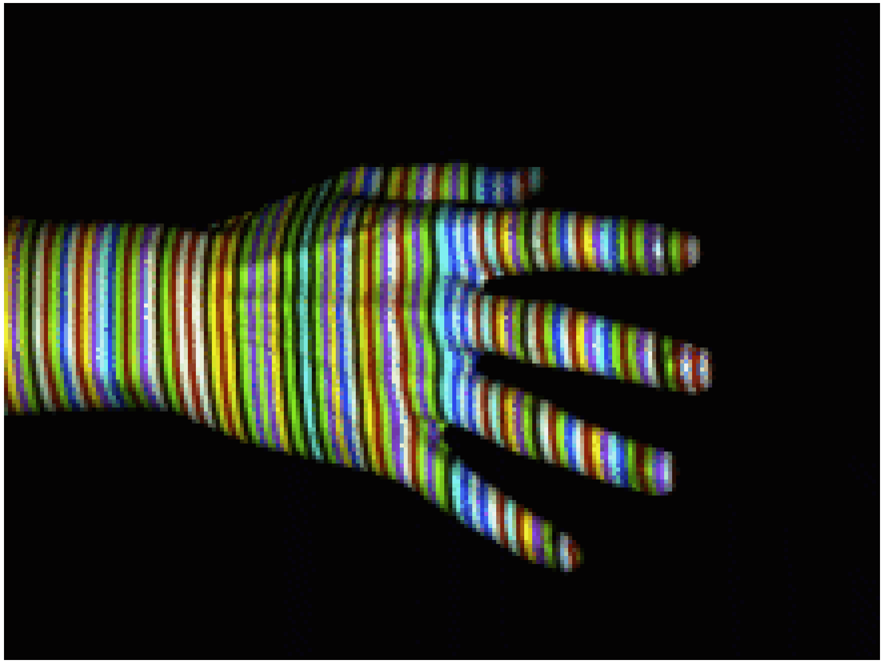


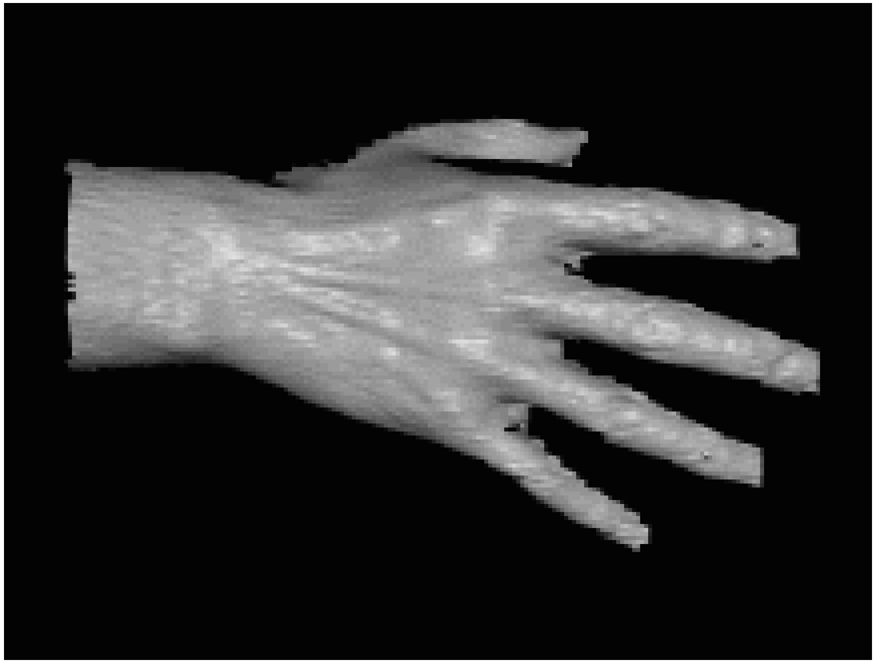
Colour De Bruijn’s sequence B(5,3)

[Zhang 2002]

STRUCTURED LIGHT: UNIQUE "SNAPSHOT" TARGET

Using a unique (« snapshot ») target reduces significantly the acquisition time and then allows to acquire mobile scenes:

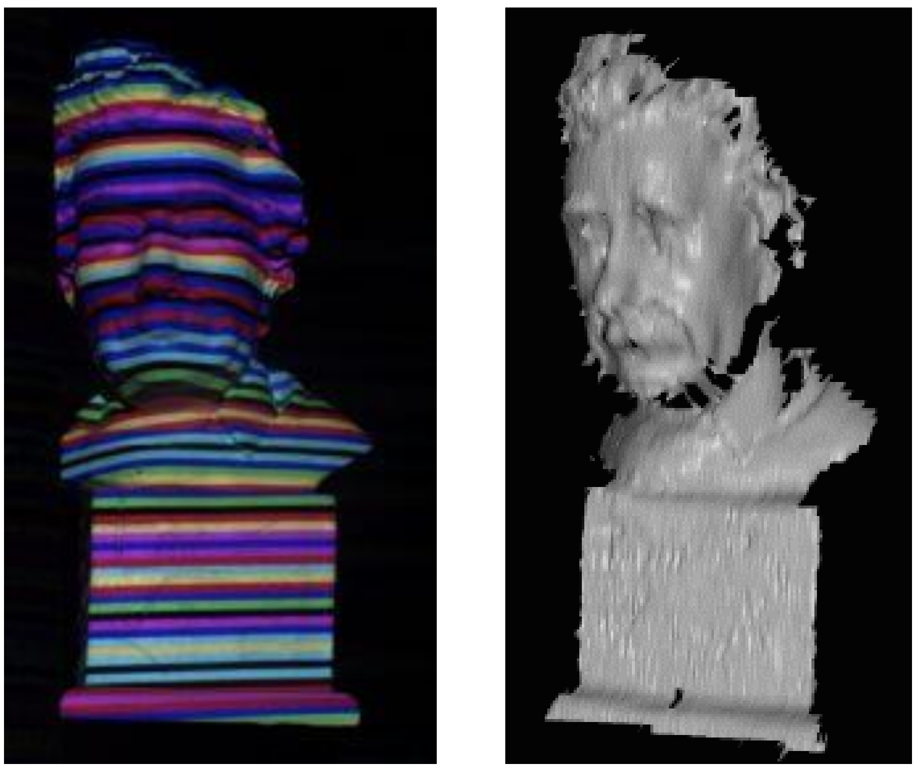




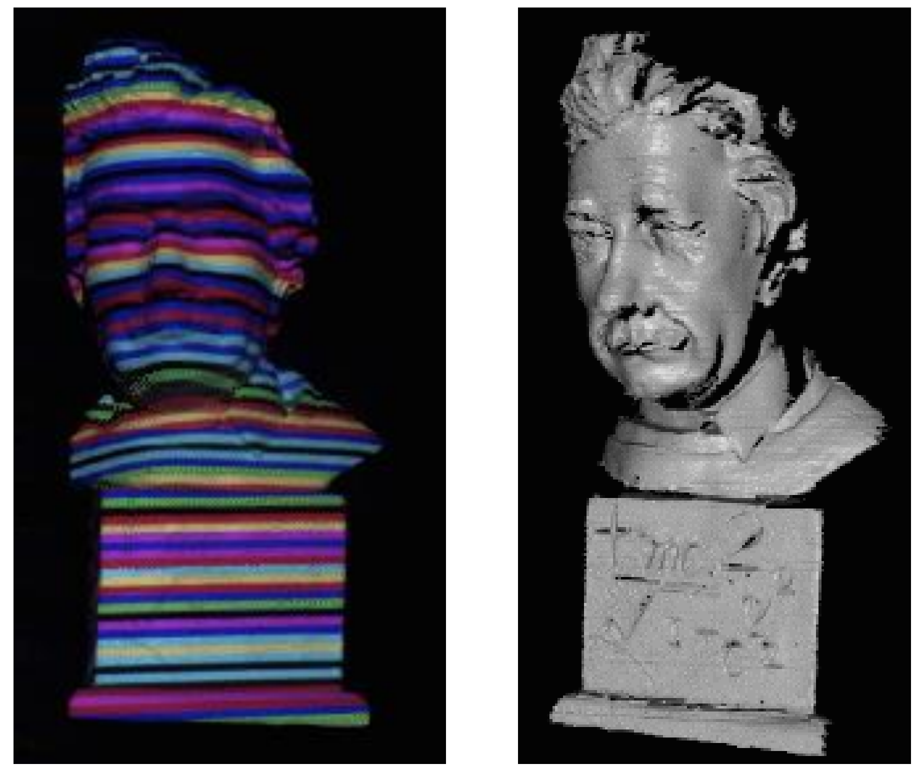
[Zhang 2002]

DE BRUIJN'S TARGET: SNAPSHOT *VS* SEQUENTIAL

Targets designed for snapshot acquisition can be used with *phase shifts* for sequential acquisitions, to improve both robustness and resolution (static scenes):



« Snapshot » acquisition

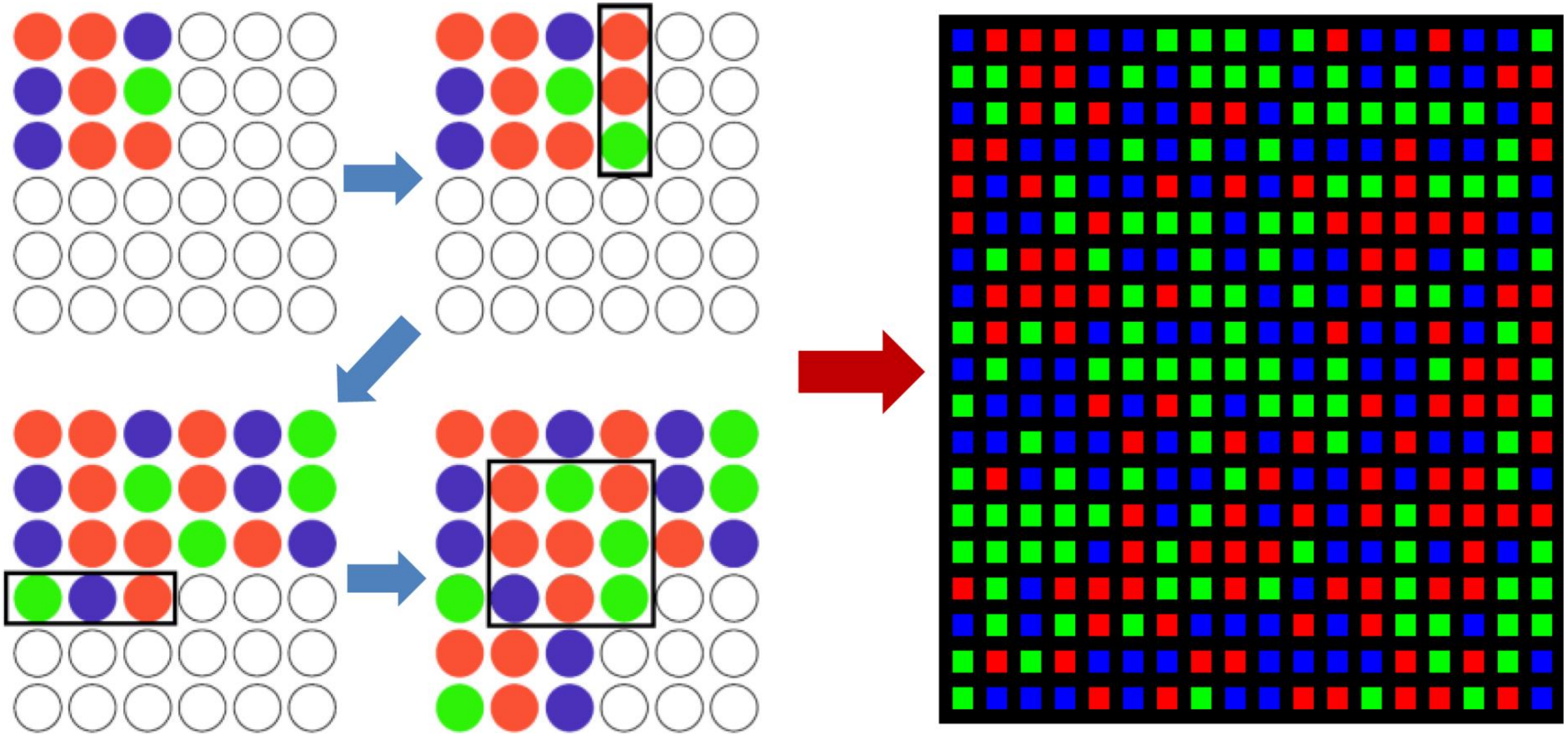


Sequential acquisition:
7 interlaced targets

[Zhang 2002]

STRUCTURED LIGHT: UNIQUE "SNAPSHOT" TARGET

2d « snapshot » target by pseudo-random patterns generated
using a brute-force algorithm:

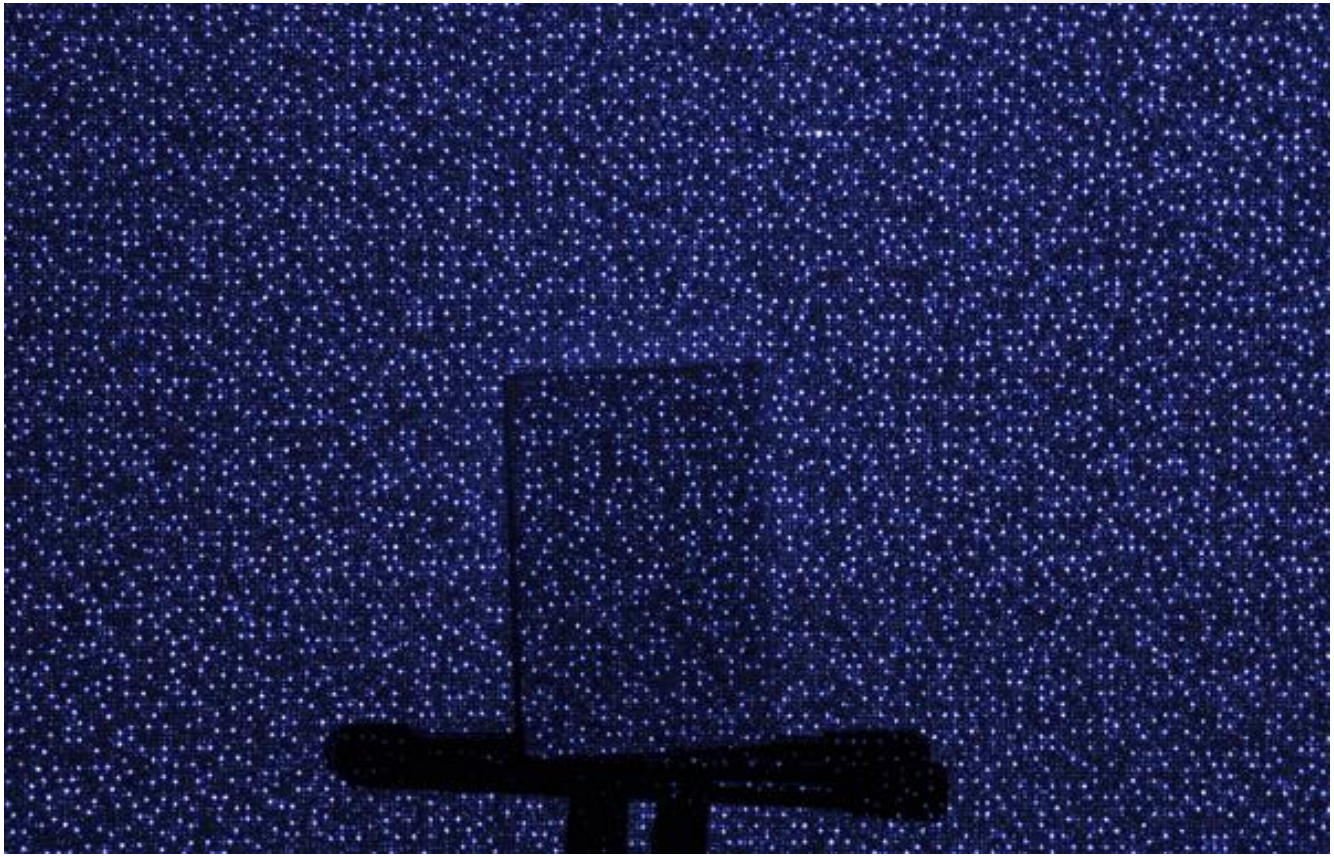


[Geng 2011]

STRUCTURED LIGHT: UNIQUE "SNAPSHOT" TARGET

The first version of the Kinect™ includes an RGB camera associated to a structured light 3d camera using a pseudo-random patterned infra-red light.

[Kinect v1 - © Microsoft]



[© futurepicture.org]

SCIENCES SORBONNE UNIVERSITÉ

Part 2: 3D CAMERAS / PASSIVE APPROACHES

For energy and / or discretion purposes, it may be better for an observation system, not to emit light.

The passive techniques get the information using only the light intensity captured by the photosensors.

The approaches presented in this chapter are all based on a non-pinhole aperture associated to a lens, by making the most of the focus and blur information:

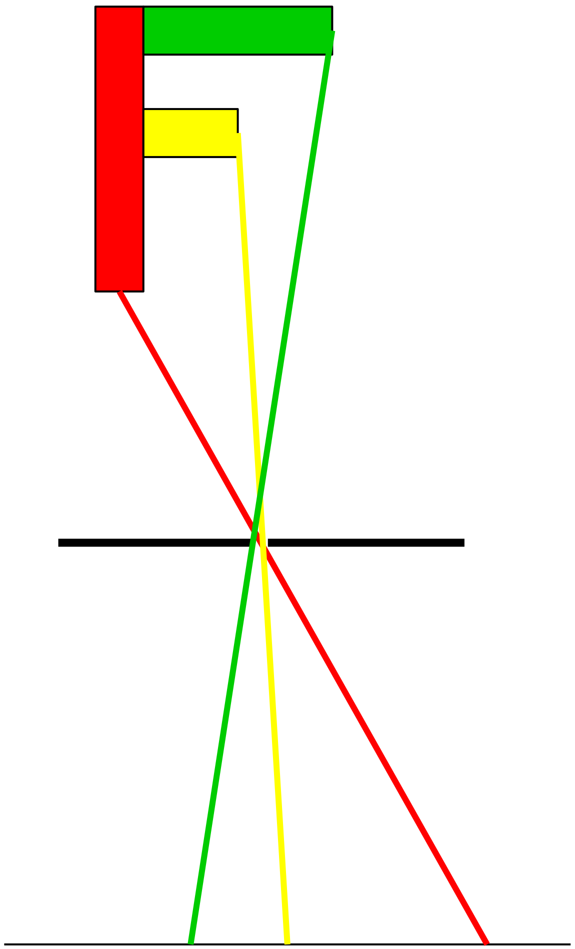
- Plenoptic camera
- Depth from (de)focus
- Coded aperture

Plenotic camera 3d Raytrix^{TM}

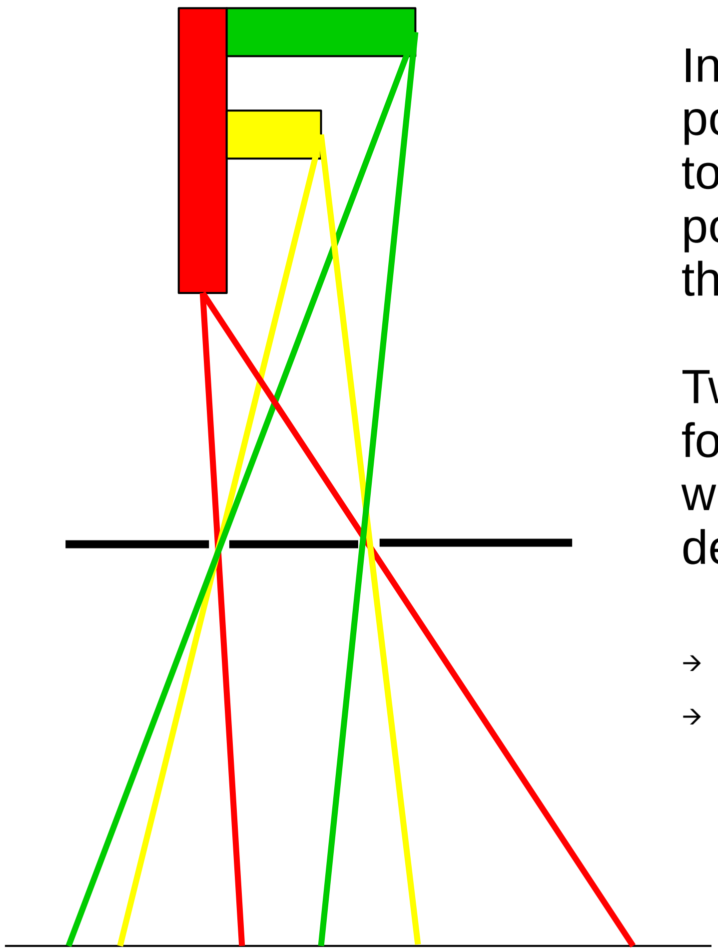
Light field plenotic camera Lytro^{TM}

SCIENCES SORBONNE UNIVERSITÉ

PASSIVE APPROACHES: PINHOLE…



(a)



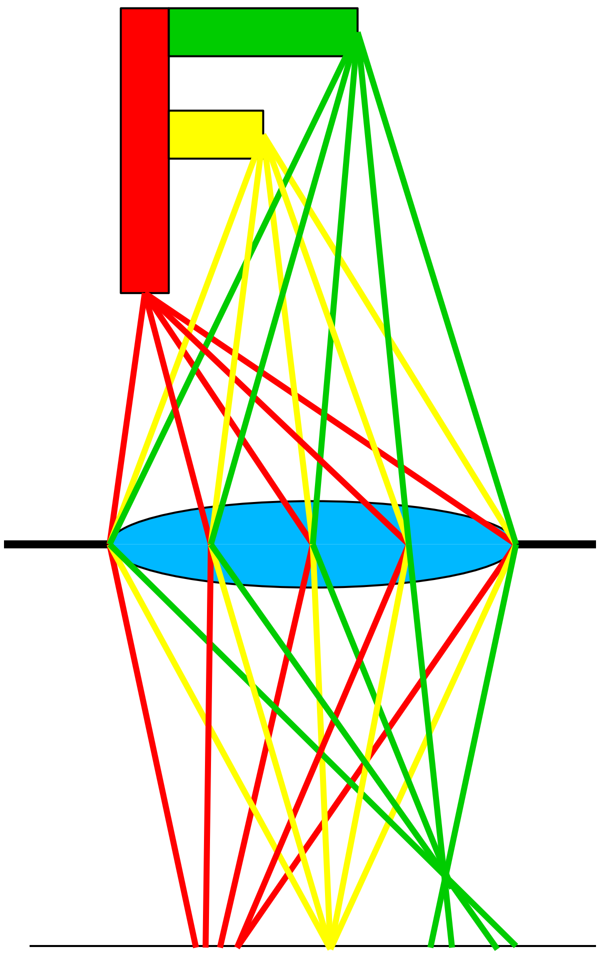
(b)

In a pinhole camera (a), every point of the image corresponds to a unique optical path. All the points appear sharp, whatever their depth.

Two distinct points of view (b), form different images, from which depth information may be deduced.

- → *Stereovision*
- → *Structure from Motion*

PASSIVE APPROACHES: …*VS* LENS



With a lens on the aperture, each point of the scene illuminates the focal plane along many different optical paths, corresponding to the line beam formed by the cone whose basis is the aperture.

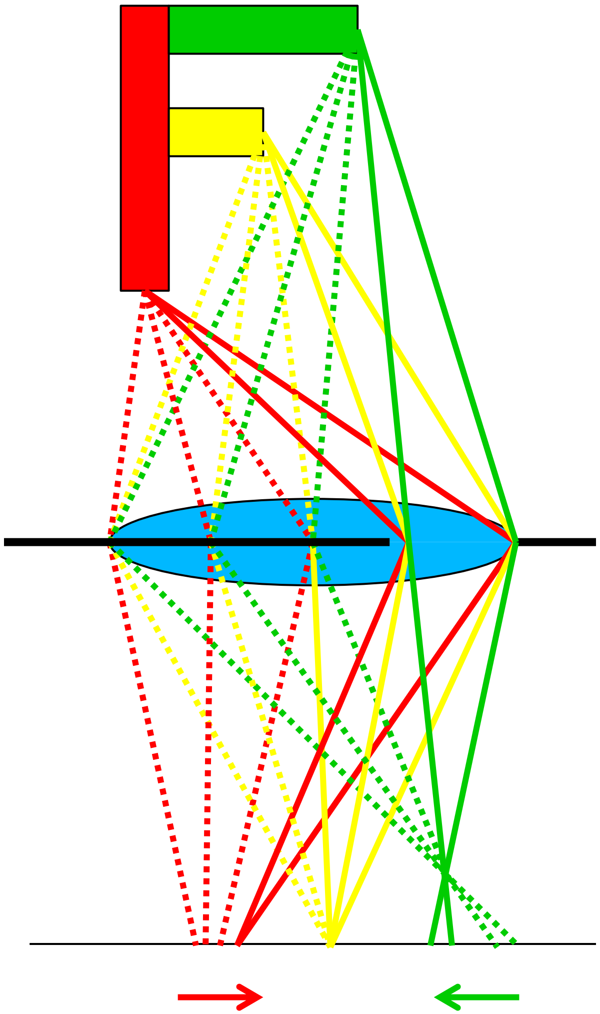
Each path line corresponds to an infinitesimal portion of the aperture, through which the scene is perceived under a particular angle.

Each infinitesimal portion then forms a pinhole-like image, and the image formed by the lens corresponds to the sum of those many « pinhole images ».

If the point is in the conjugate plane of the focal plane (sharpness plane), all the different paths converge on the image, and the point appears sharp, otherwise it appears more or less blurred depending on its distance to the sharpness plane.

→ *Depth from (de)focus*

PASSIVE APPROACHES: LENS AND APERTURES



By using an excentric aperture (figure), a sub-set of the optical paths is selected, reducing both the blur and the light intensity.

Points in the sharpness plane (yellow lines) remain at the same location in the image plane.

Closer points (**red lines**) are deviated in the direction of the aperture.

Further points (**green lines**) are deviated in the inverse direction.

→ *Coded aperture:*
Modify the geometry of the aperture for an easier interpretation of the blur (≈ point spread function of the aperture).

→ *Plenoptic camera:*
Separate physically the different optical paths within sub-beams focalised on distinct parts of the sensor.

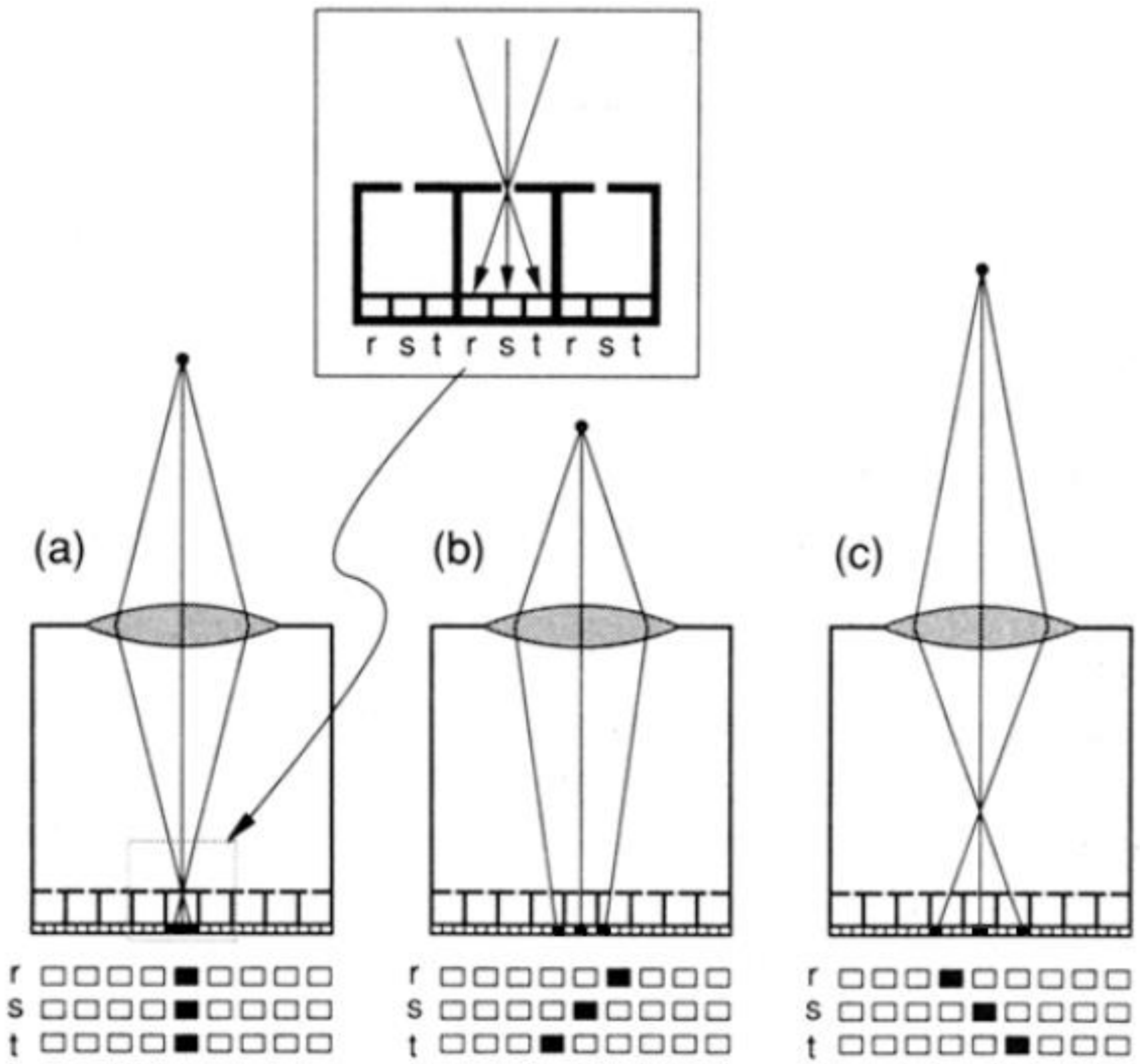
SCIENCES SORBONNE UNIVERSITÉ

PASSIVE 3D: PLENOPTIC CAMERA

In a plenoptic camera, the optical paths are separated within sub-beams that are focalised on different parts of the sensor.
(Figure: mini-pinholes, but also 1d lenticular grid, or 2d micro-lens grid).

The captured information is then composed of one macro-image made of many hyper-pixels (or micro-images).
(See figure:
- Macro-image of 1x9 hyper-pixels.
- Hyper-pixel of size 1x3.)

The plenoptic image then captures a 4d information: $I(x,y,\xi,\varsigma)$, where (x,y) is the direction of a point illuminating the aperture (light cone), and (ξ,ς) a particular view of this point through the aperture.

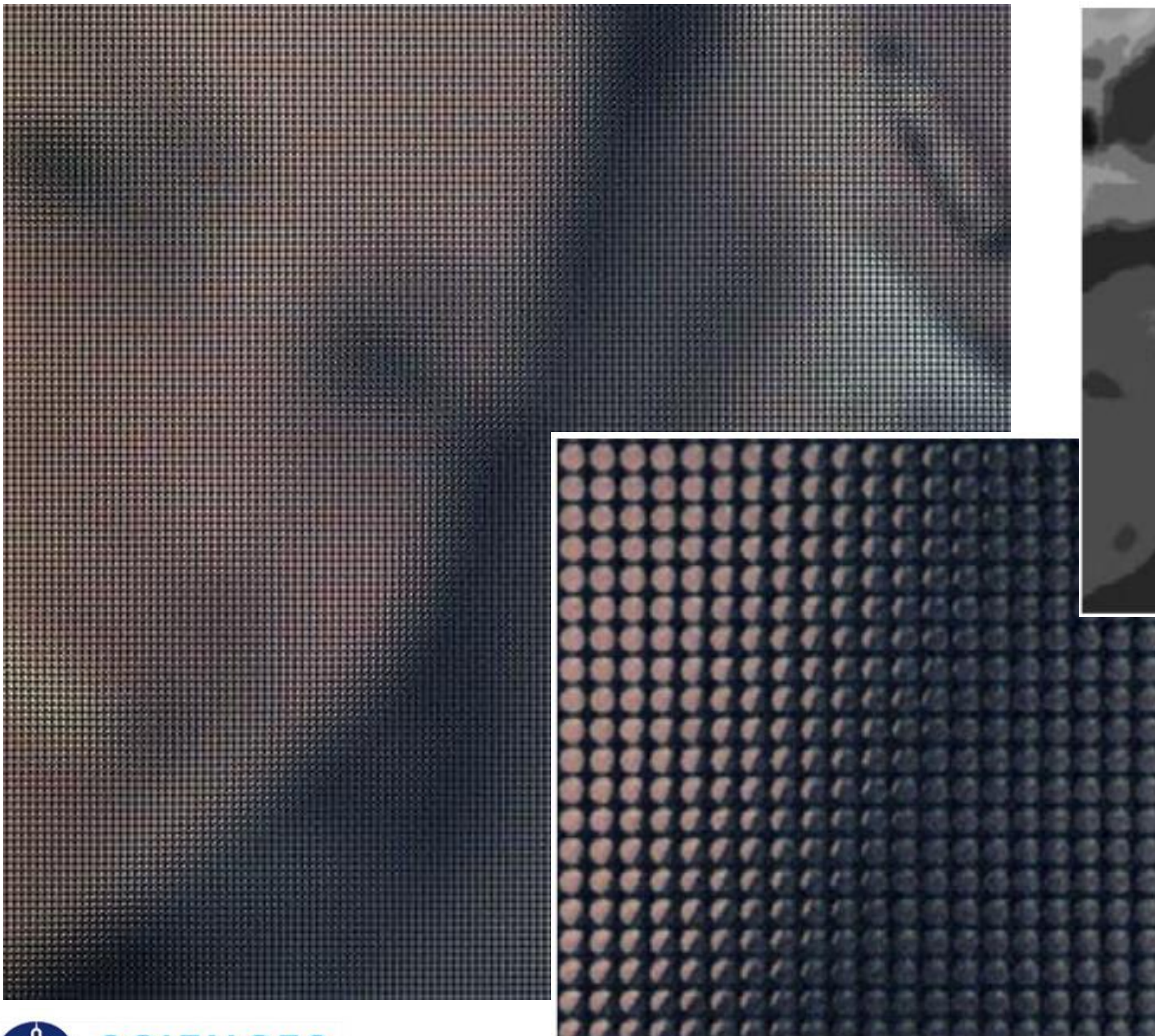


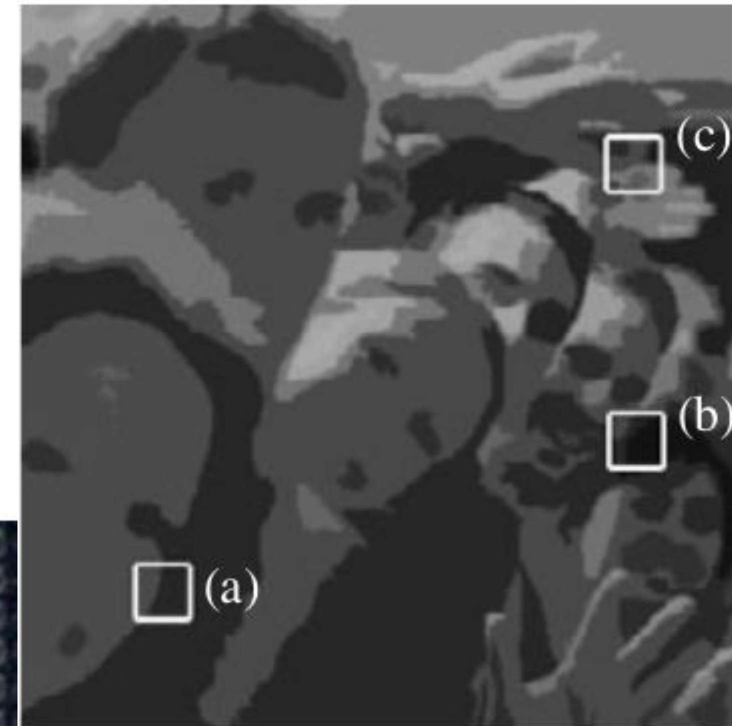
[Adelson 1992]

PLENOPTIC CAMERA: MACRO-IMAGE

[Ng 2005]

PLENOPTIC CAMERA: MICRO-IMAGES



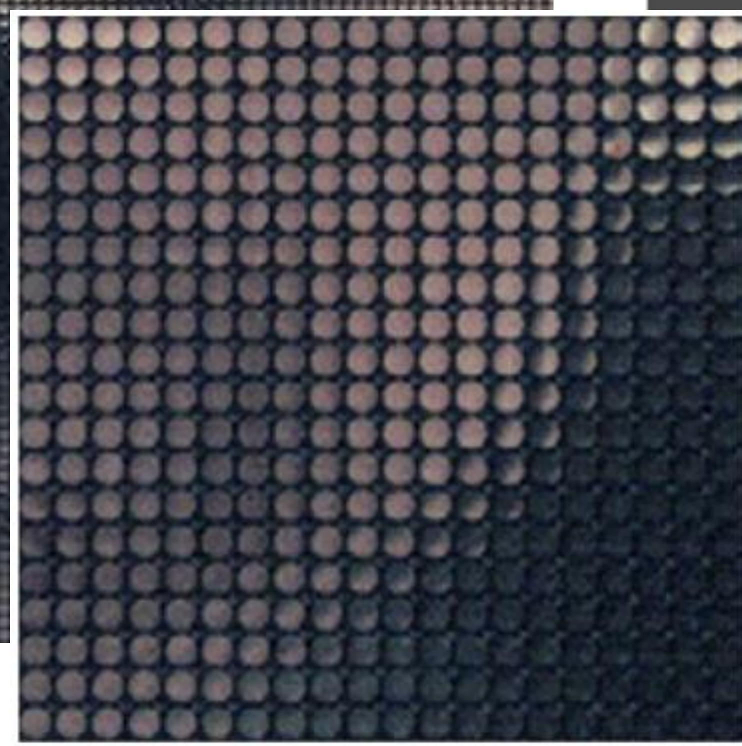


(a)

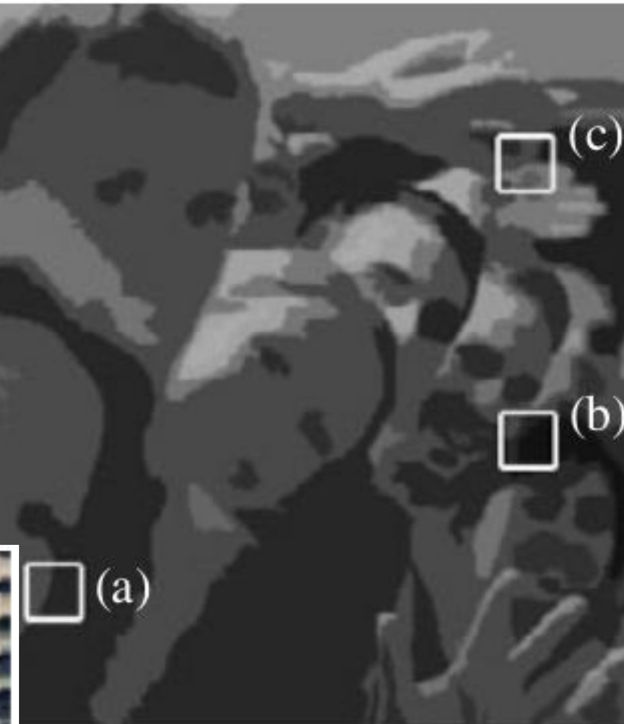
Micro-images (hyper-pixels) of points that are closer than the sharpness plane *have a contrast in the same sense* as the macro-image.

[Ng 2005]

PLENOPTIC CAMERA: MICRO-IMAGES



(b)

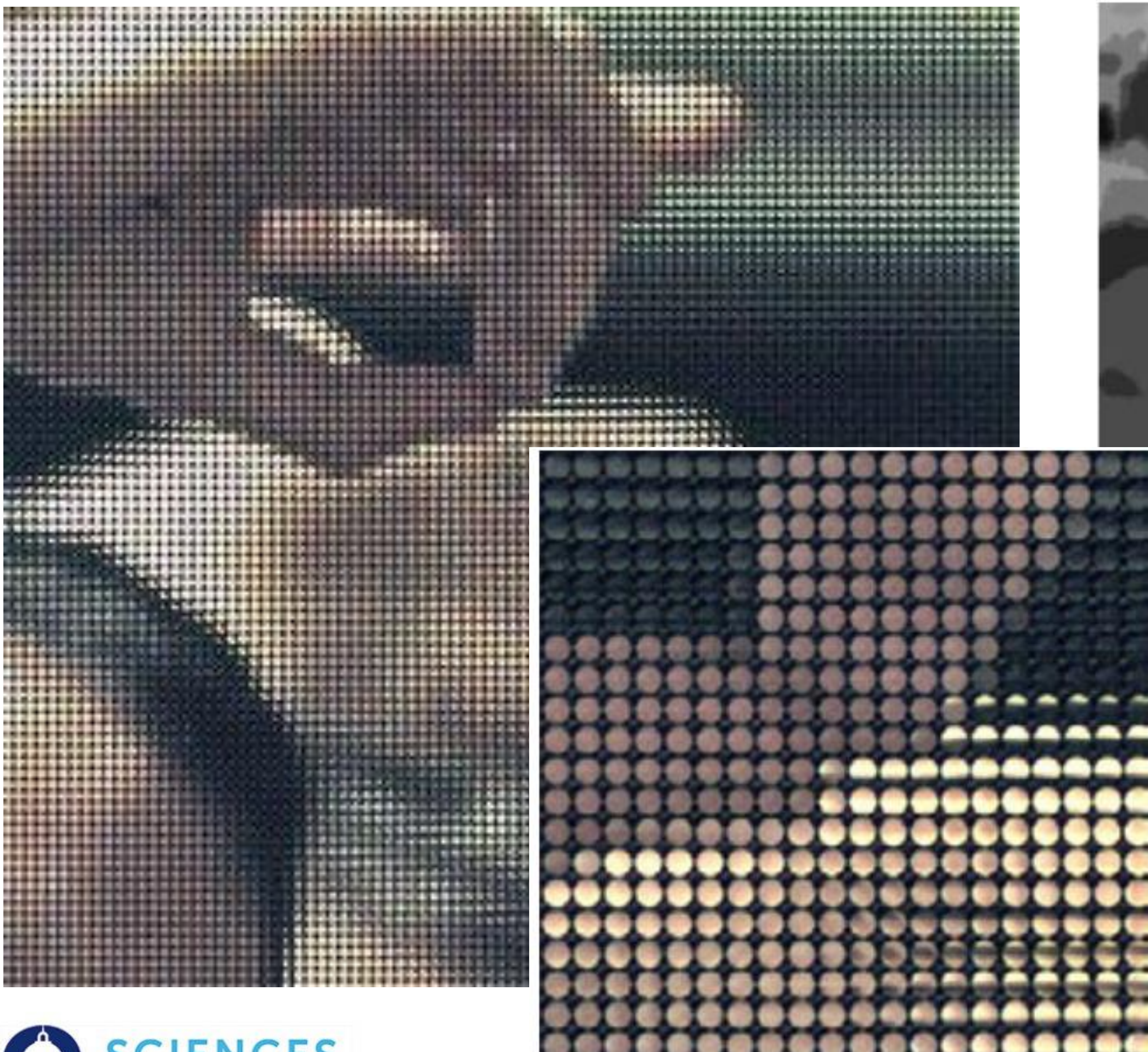


Micro-images (hyper-pixels) of points that are further than the sharpness plane *have an inverted contrast* with respect to the macro-image.

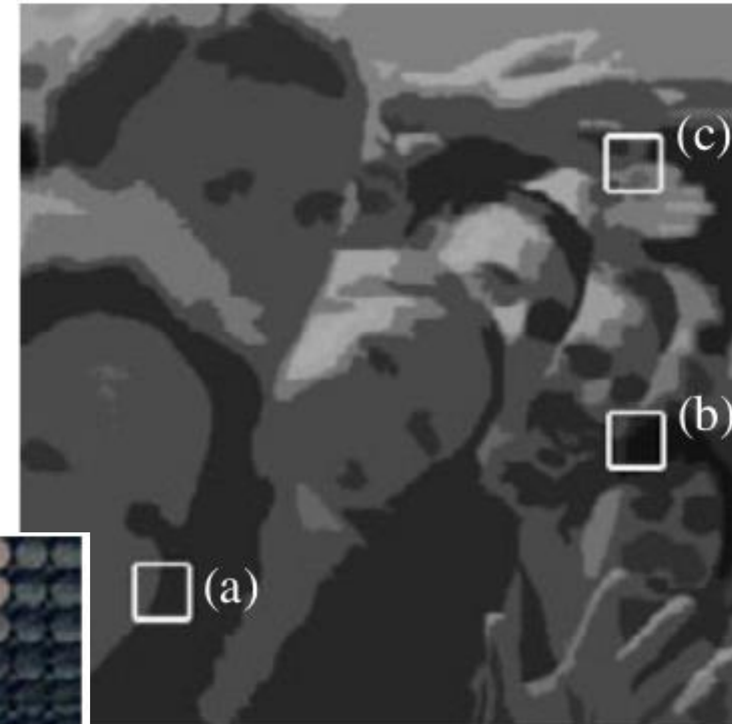
[Ng 2005]

SCIENCES SORBONNE UNIVERSITÉ

PLENOPTIC CAMERA: MICRO-IMAGES



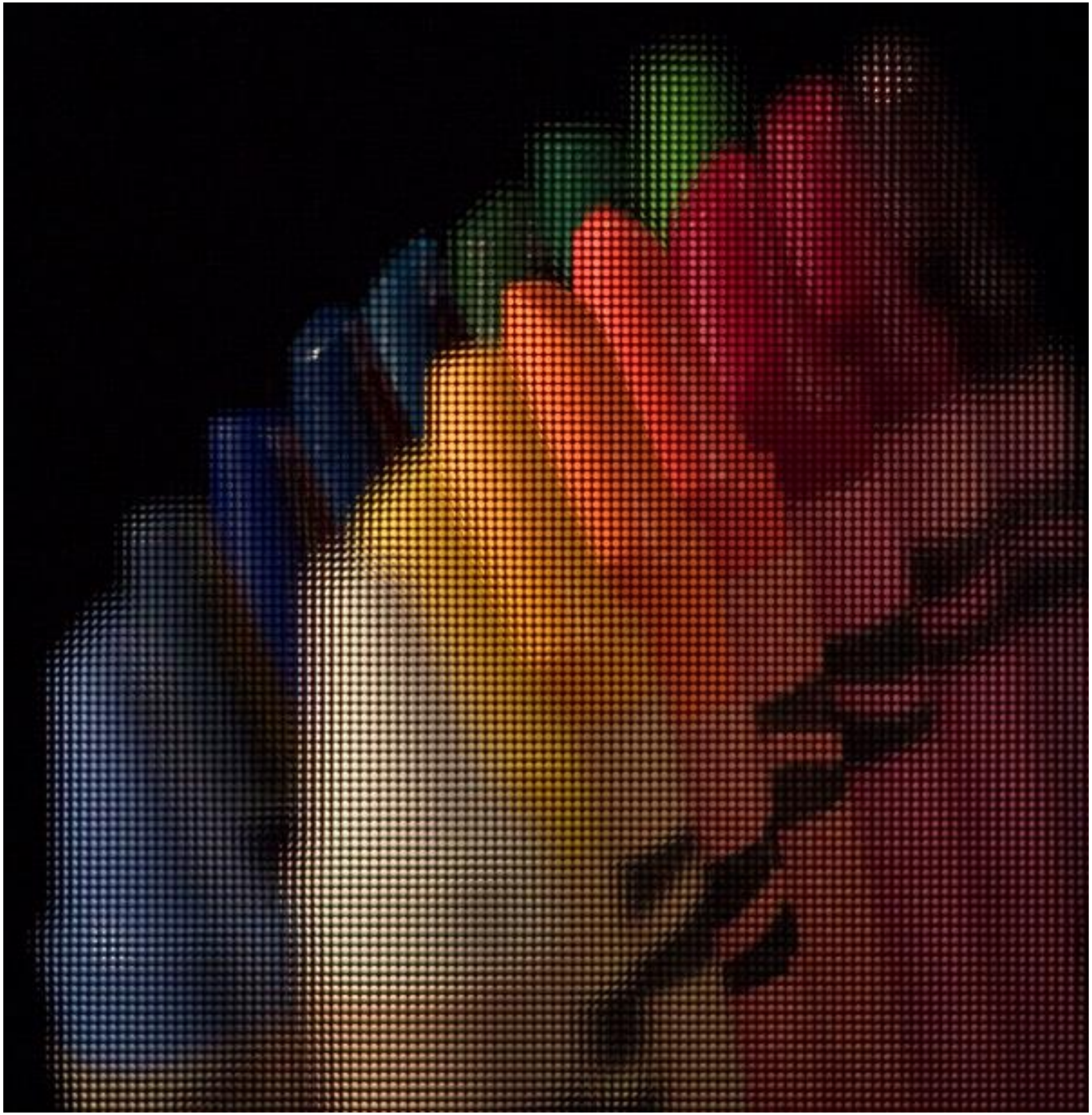
(c)



Micro-images (hyper-pixels) of points in the sharpness plane *form homogeneous regions.*

[Ng 2005]

PLENOPTIC CAMERA: MACRO-IMAGE

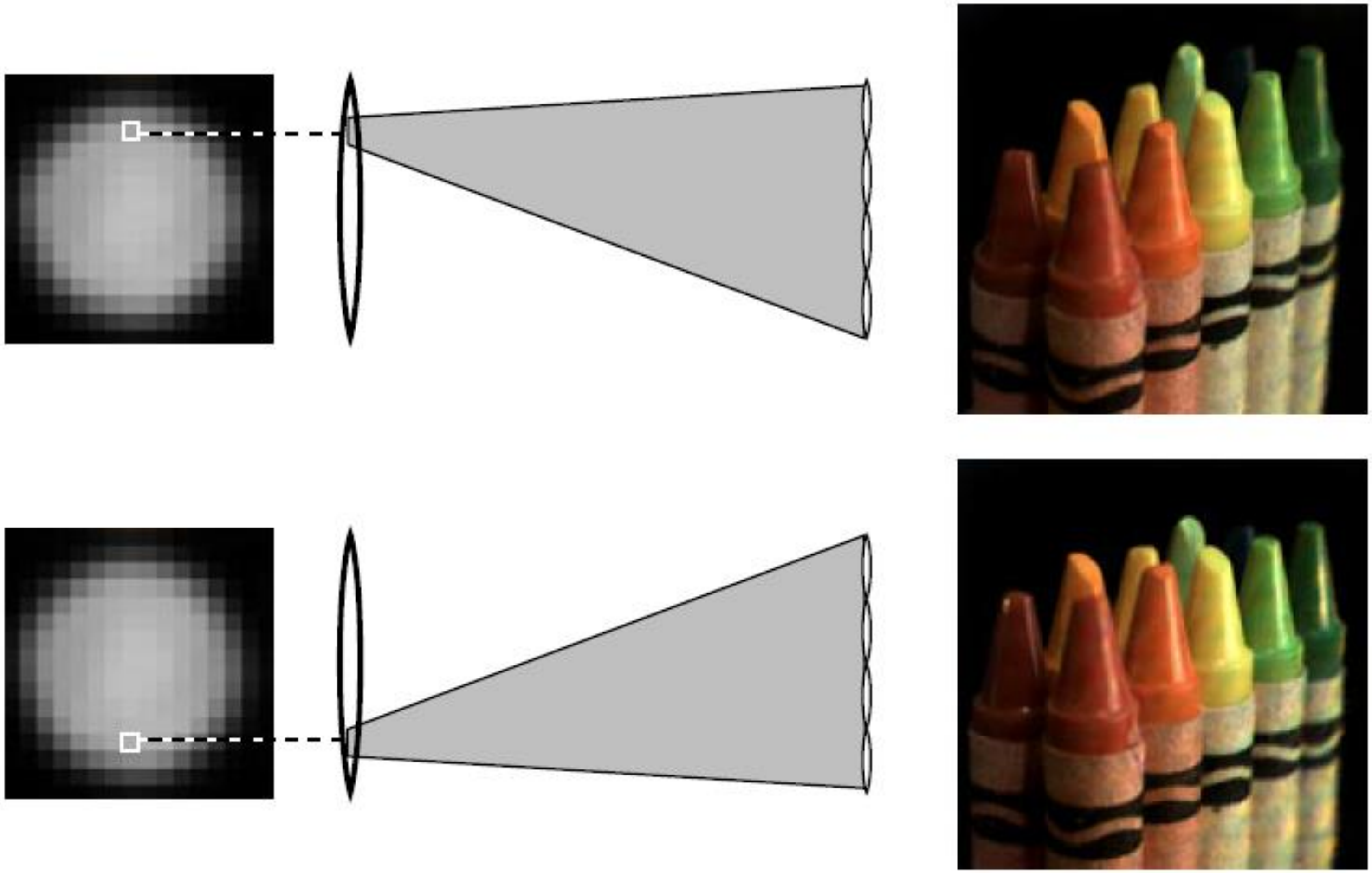


[Ng 2005]

PLENOPTIC: MACRO-IMAGE AND DUAL MACRO-IMAGES

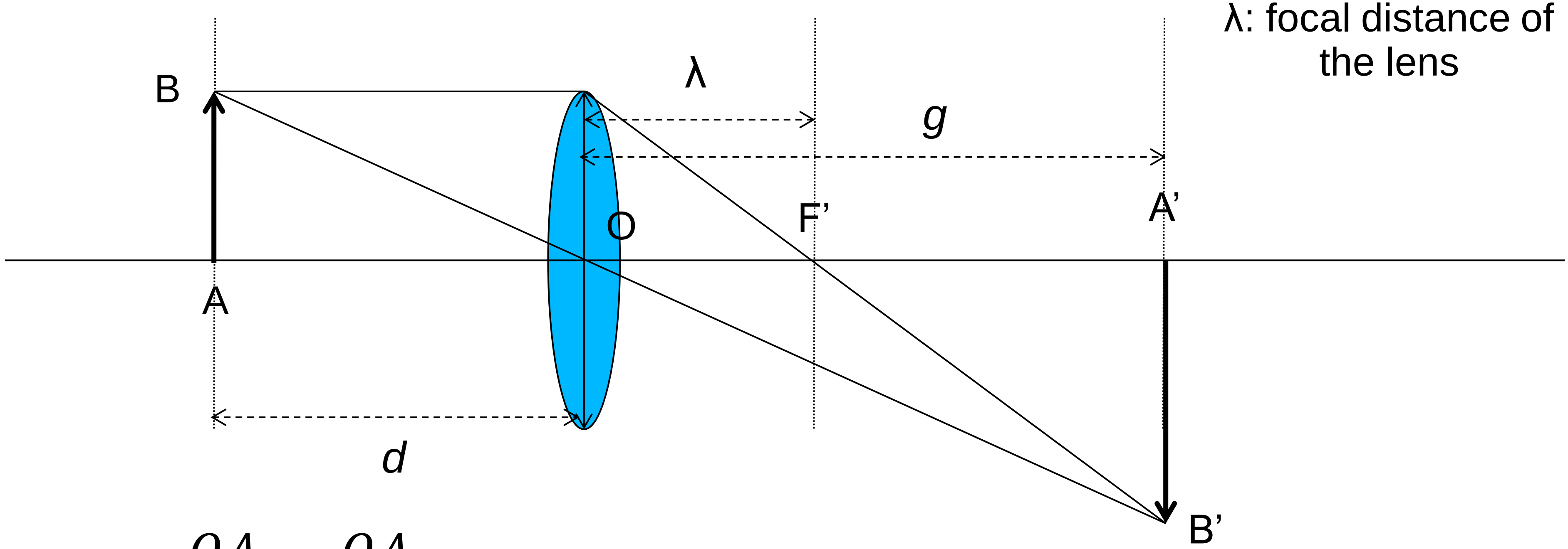
Dual macro-images are made by recomposing *m* x *m* sub-sampled images of size *n* x *n* from the homologous pixels of all the micro-images, where *n* x *n* is the number of micro-images (resolution of the macro-image), and *m* x *m* is the resolution of the micro-image.

Dual macro-images then correspond to a partition of the aperture into distincts viewpoints and then present parallax differences, from which depth information can be deduced by matching (*single-lens stereo*).



[Ng 2005]

GEOMETRY OF THE THIN CONVERGENT LENS



λ: focal distance of the lens

$$\frac{OA}{OF'} = \frac{OA}{OA'} + 1$$

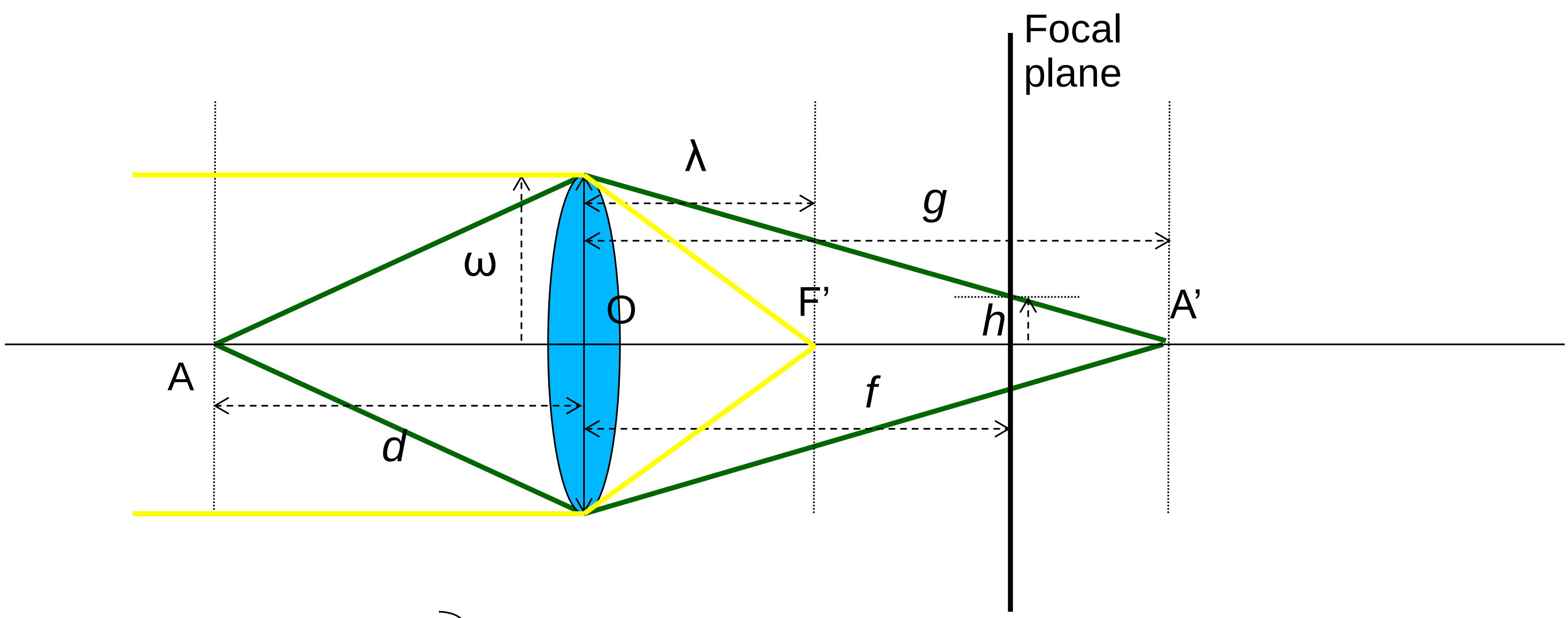
and then:

$$\frac{1}{d} + \frac{1}{g} = \frac{1}{\lambda}$$

Thin lens equation

- *d*: distance of point B to the aperture plane (depth)
- *g* : distance between the aperture and the focalisation plane of point B

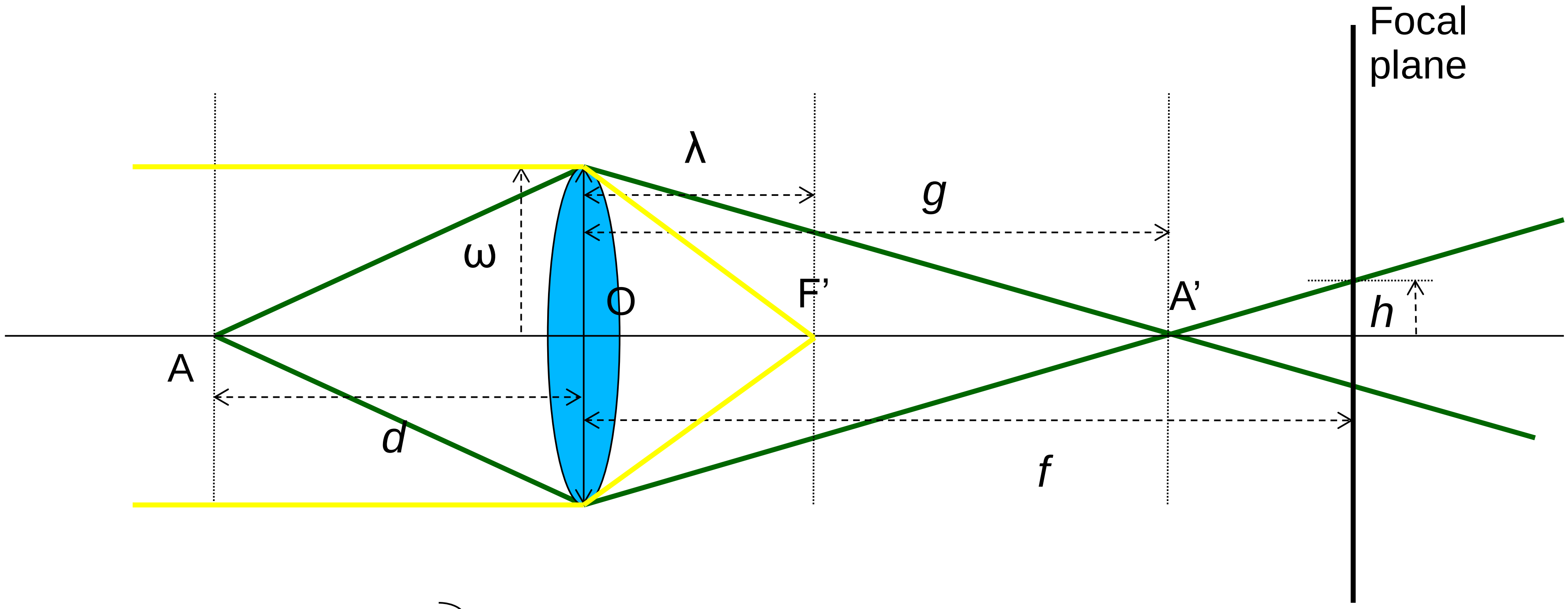
RELATION FOCUS / DISTANCE: SHORT FOCAL



$$\left.\begin{aligned}\frac{f}{g} &= 1 - \frac{h}{\omega}\\ \frac{1}{d} + \frac{1}{g} &= \frac{1}{\lambda}\end{aligned}\right\} \quad \frac{1}{d} = \left(\frac{1}{\lambda} - \frac{1}{f}\right) + \frac{h}{f\omega}$$

λ: lens focal
ω: aperture
f: image focal
h: defocus width

RELATION FOCUS / DISTANCE: LONG FOCAL



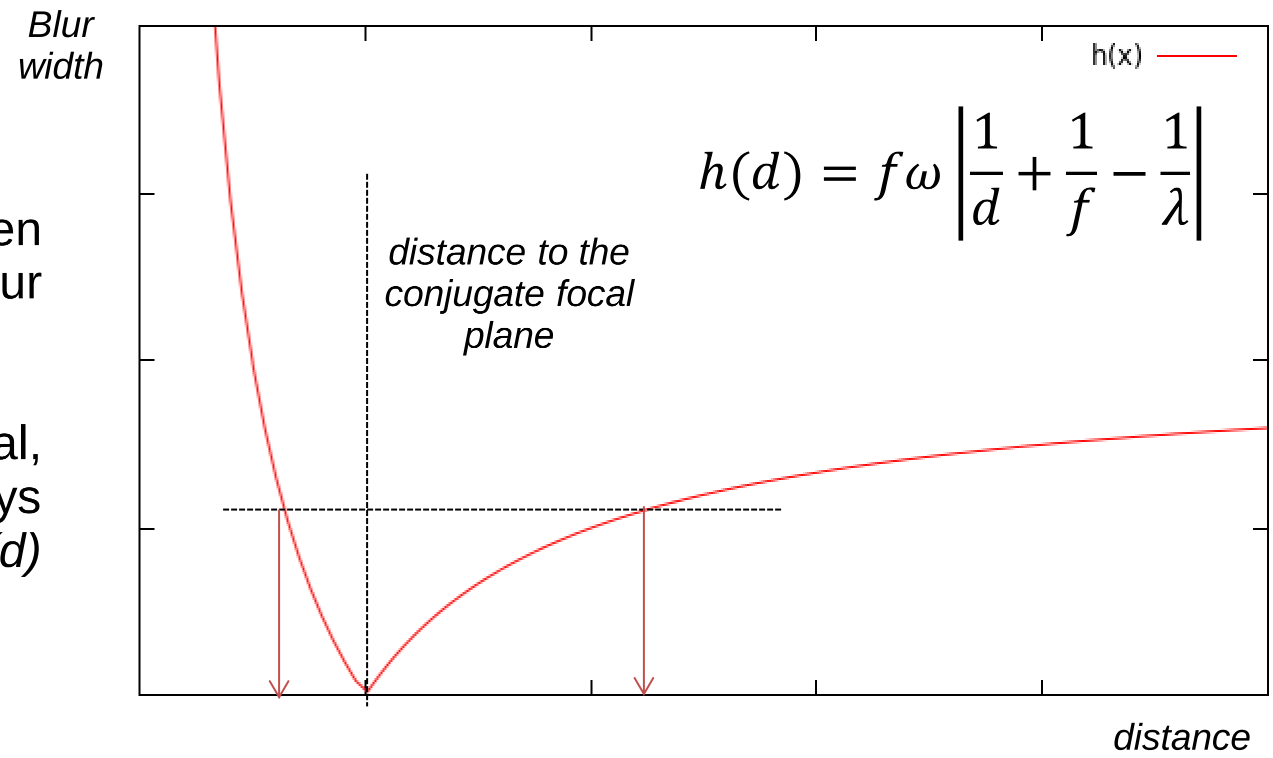
$$\left.\begin{array}{l} \dfrac{f}{g} = 1 + \dfrac{h}{\omega} \\[2ex] \dfrac{1}{\lambda} = \dfrac{1}{d} + \dfrac{1}{g} \end{array}\right\} \quad \frac{1}{d} = \left(\frac{1}{\lambda} - \frac{1}{f}\right) - \frac{h}{f\omega}$$

λ: lens focal
ω: aperture
f: image focal
h: defocus width

PASSIVE 3D: DEPTH FROM (DE)FOCUS

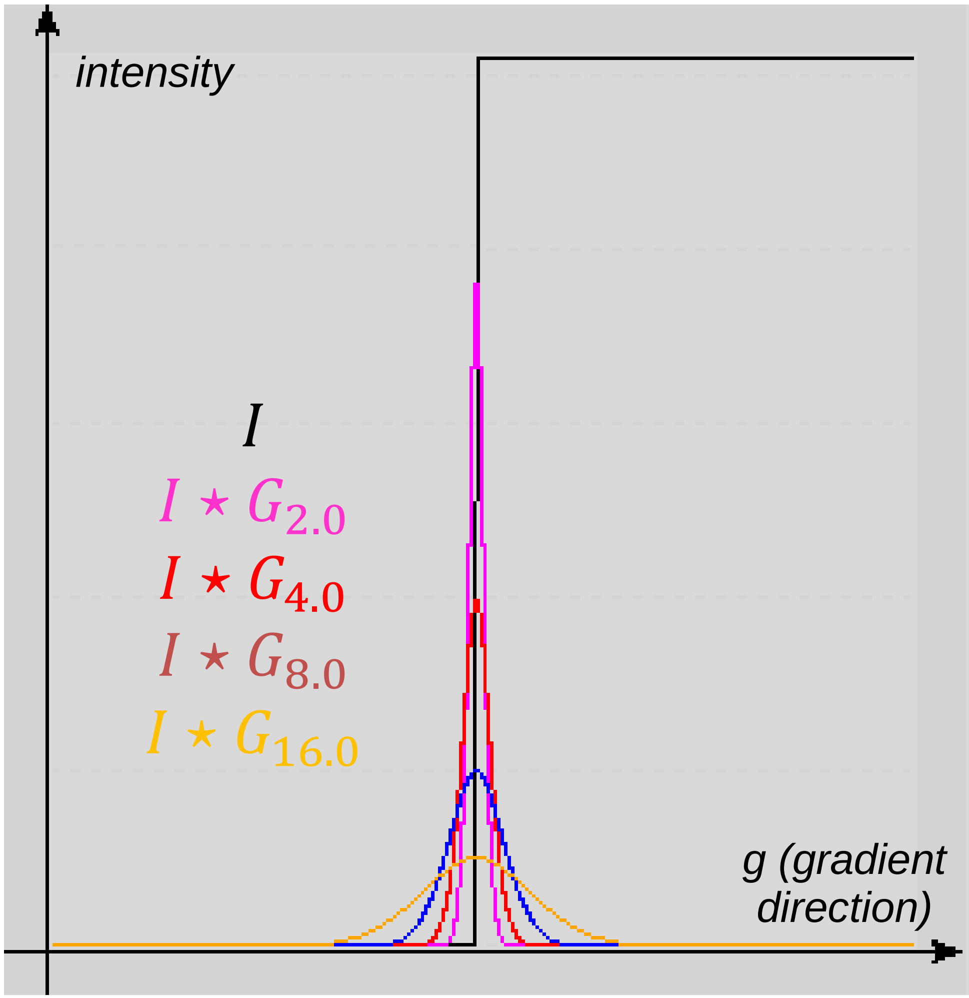
Estimating the distance can then be made by estimating the blur width in the image.

Without prior on the image focal, a single measure is always ambiguous, the function *h(d)* being not injective (Figure).



To perform direct measurement of the blur width by image processing, an hypothesis on the structure of the sharp image is necessary: impulsion, step-like contour, in order to predict the effect of blur on this structure.

PASSIVE 3D: DEPTH FROM (DE)FOCUS



If the blur is modelled by a 2d Gaussian convolution whose standard deviation depends on *h*, *h* can be deduced from the effect of blur on a step-like contour structure, by measuring the local maximum of the gradient value in the direction orthogonal to the step.

Those structures correspond to the classic definition of contours, i.e. the zero-crossings of the second derivative in the gradient direction *g*:

$$C_I = \left\{ x; \frac{\partial^2 I}{\partial g^2}(x) = 0 \right\}$$

Question: how to justify the use of a Gaussian blur model when the geometric optics predicts a gate (square) function?

PASSIVE 3D: DEPTH FROM (DE)FOCUS

$$I(x) = (I^H(x), I^S(x), I^V(x))$$

$\frac{\partial I^V}{\partial g}(x)$ (gradient magnitude)

PASSIVE 3D: DEPTH FROM (DE)FOCUS

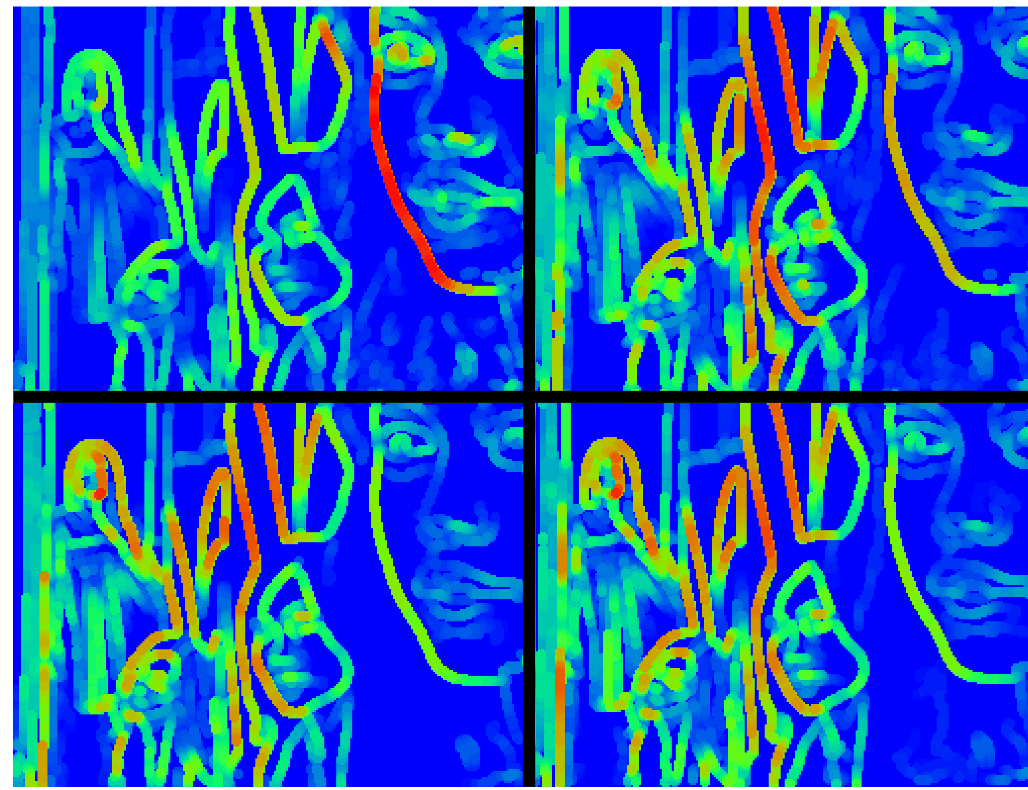
$I(x)$

$C_I = \left\{ x; \frac{\partial^2 I^V}{\partial g^2}(x) = 0 \right\}$ (contours)

PASSIVE 3D: DEPTH FROM (DE)FOCUS

Measuring the gradient magnitude along the contours allows estimating the blur width h, but remains ambiguous regarding the position with respect to the sharpness plane.

$I(x)$



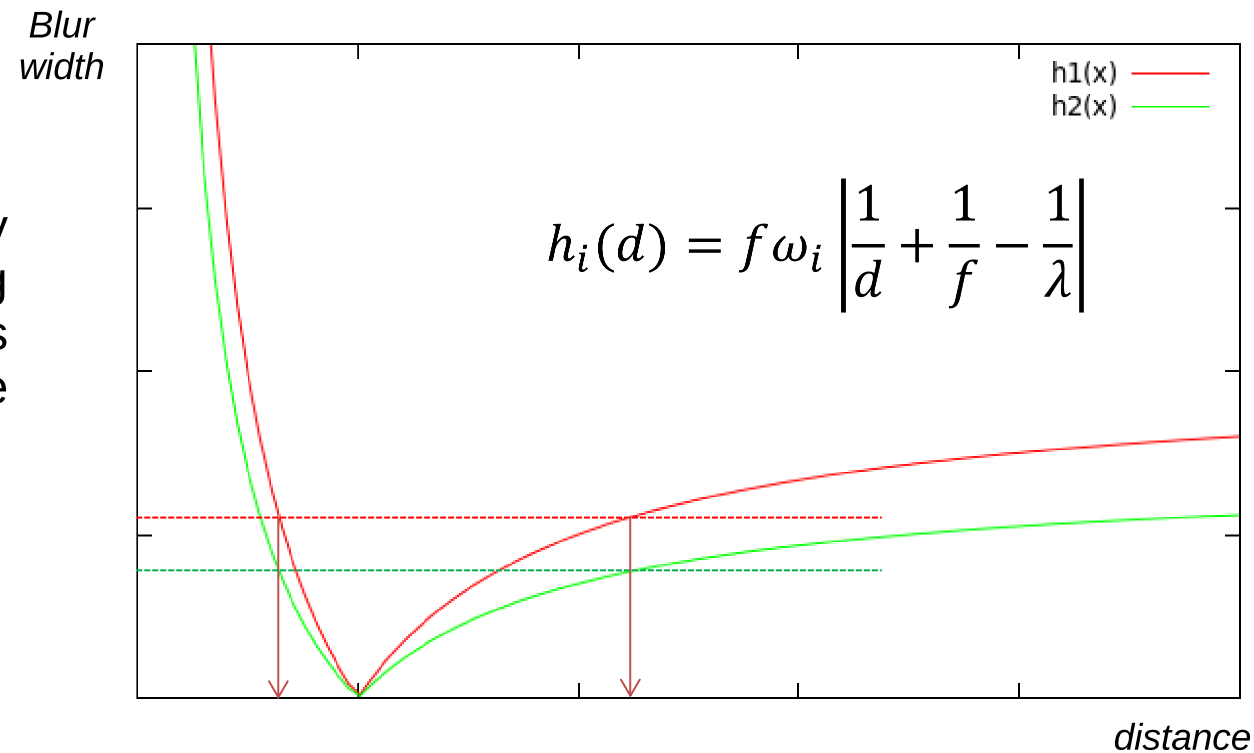
Measuring the blur width along the contours

[Pentland 1987]

Idea: repeat the measure while varying the aperture ω and/or the image focal f ?

PASSIVE 3D: DEPTH FROM (DE)FOCUS

The blur width depends linearly of the aperture, then using different apertures only does not disambiguate the distance from the sharpness plane:



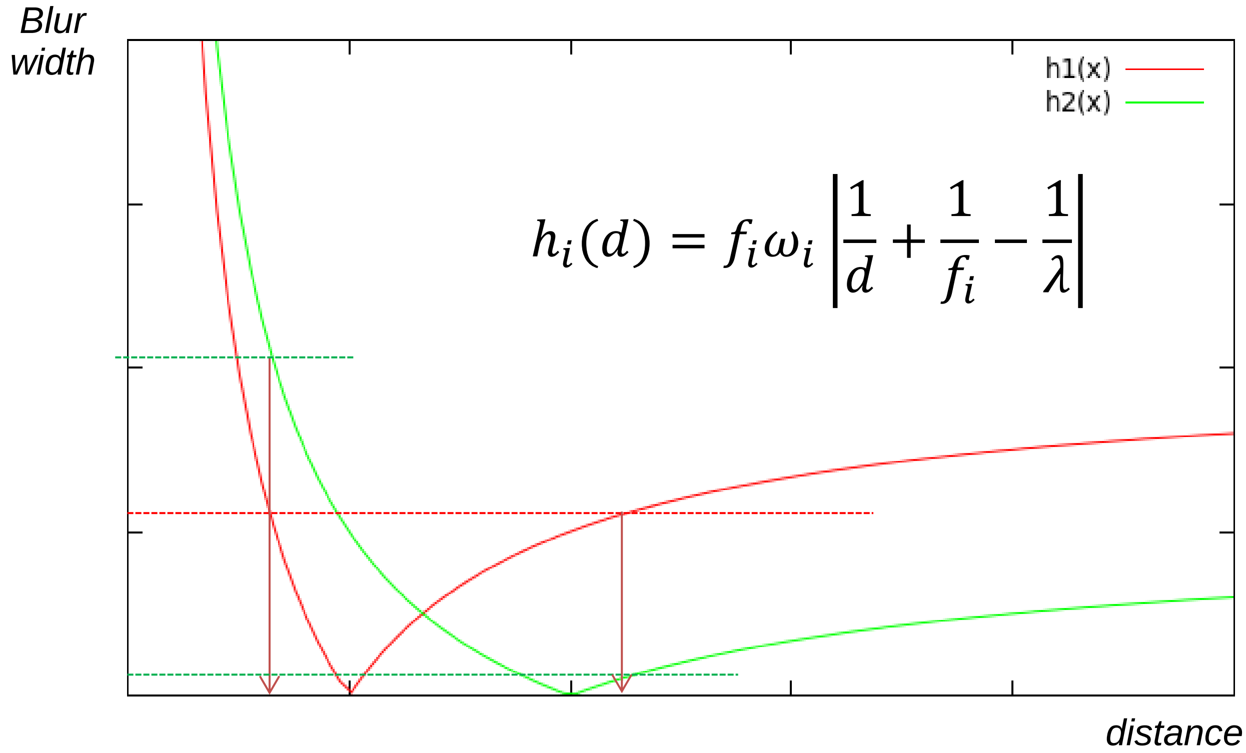
Constant focal, variable aperture.

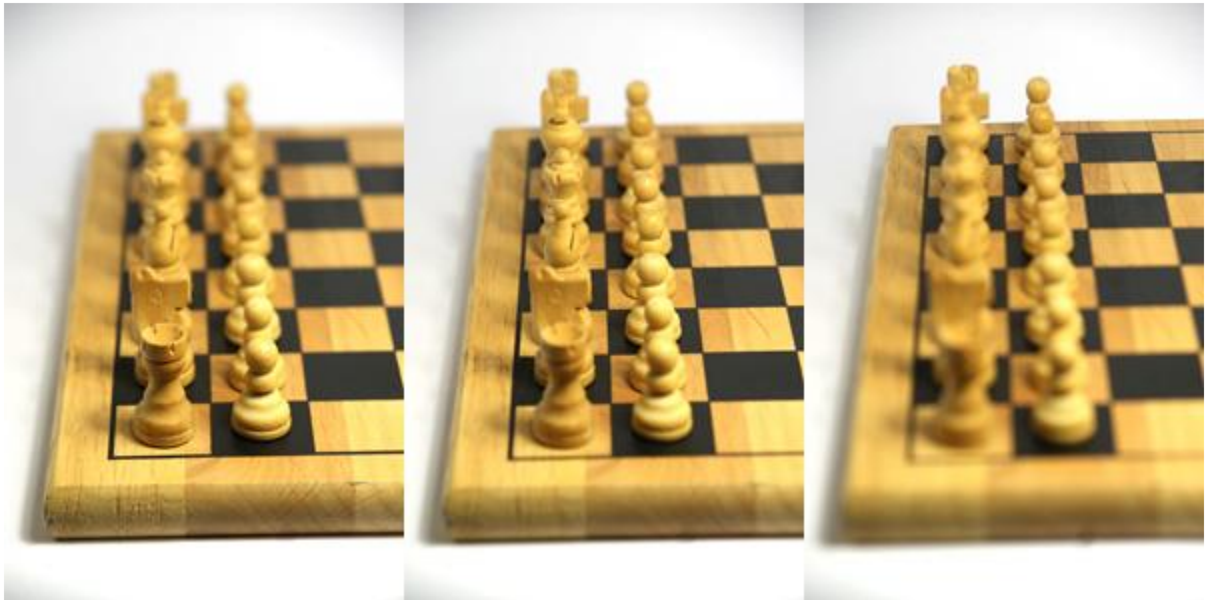
PASSIVE 3D: DEPTH FROM (DE)FOCUS

In contrast, using several couples (aperture, image focal) allows to deduce the distance from the blur width in an absolute manner.

(Figure: product $f_i \omega_i$ constant)

[Pentland 1987]





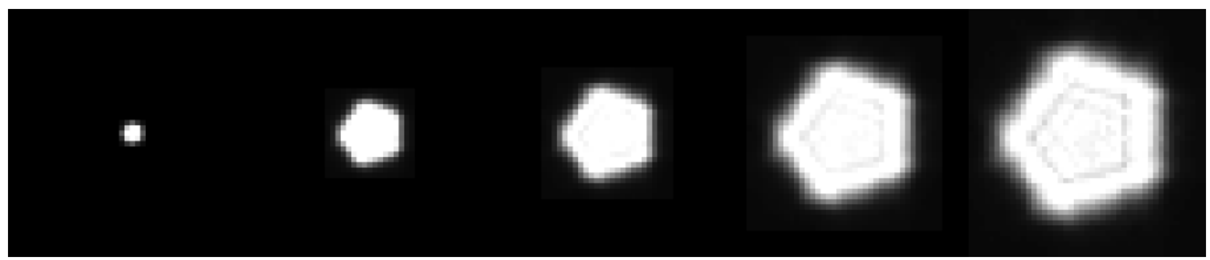
Constant aperture, variable focal.

BLUR MODEL *VS* APERTURE CALIBRATION

The Gaussian kernel is considered a better blur model than the gate function because the blur is actually the combination of several phenomena: diffraction, chromatic aberrations, discretisation, that lead to the composition of several convolutions.

However a better alternative to blur models is to perform an aperture calibration of the camera by recording the different images formed by one point for different focalisation distances (point spread functions of the convolution kernels).

[Levin 2007]



Traditional 5-blade diaphragm and the family $\{g_d\}_{d \in D}$ *of calibrated kernels.*

Estimating the right distance is then equivalent to finding the kernel g_d which best corresponds to the local observation.

The « direct » estimation being only possible on contours, indirect estimation is used instead, using deconvolution…

BLUR ESTIMATION BY DECONVOLUTION

I the observed image

$\{g_d\}_{d \in D}$ the family of calibrated convolution kernels, indexed by distance

J_d the deconvolution of I by g_d

The reconstruction error $\varepsilon_d(x)$ at pixel x and distance d is defined as:

$$\varepsilon_d(x) = \sum_{y \in W_x} \|I - J_d \star g_d\|^2$$

where W_x is a spatial neighbourhood of x.

Distance estimation is then performed as follows:

$$d_{opt}(x) = \arg\min_{d \in D} \varepsilon_d(x)$$

DECONVOLUTION: INVERSE AND WIENER FILTERING

The problem is now equivalent to image deconvolution (restoration), the convolution kernel at the origin of the blur being known (non-blind).

Quick sketch of non-blind deconvolution:

$$F = I \star g_d \xrightarrow{\textit{Fourier transform}} \tilde{F} = \tilde{I} \times \widetilde{g_d} \xrightarrow{\textit{Inverse filter}} \tilde{J_d} = \frac{\tilde{F}}{\widetilde{g_d}} \xrightarrow{\textit{Inverse Fourier transform}} J_d$$

Not usable because of the zeros of $\widetilde{g_d}$ and additive noise!!!

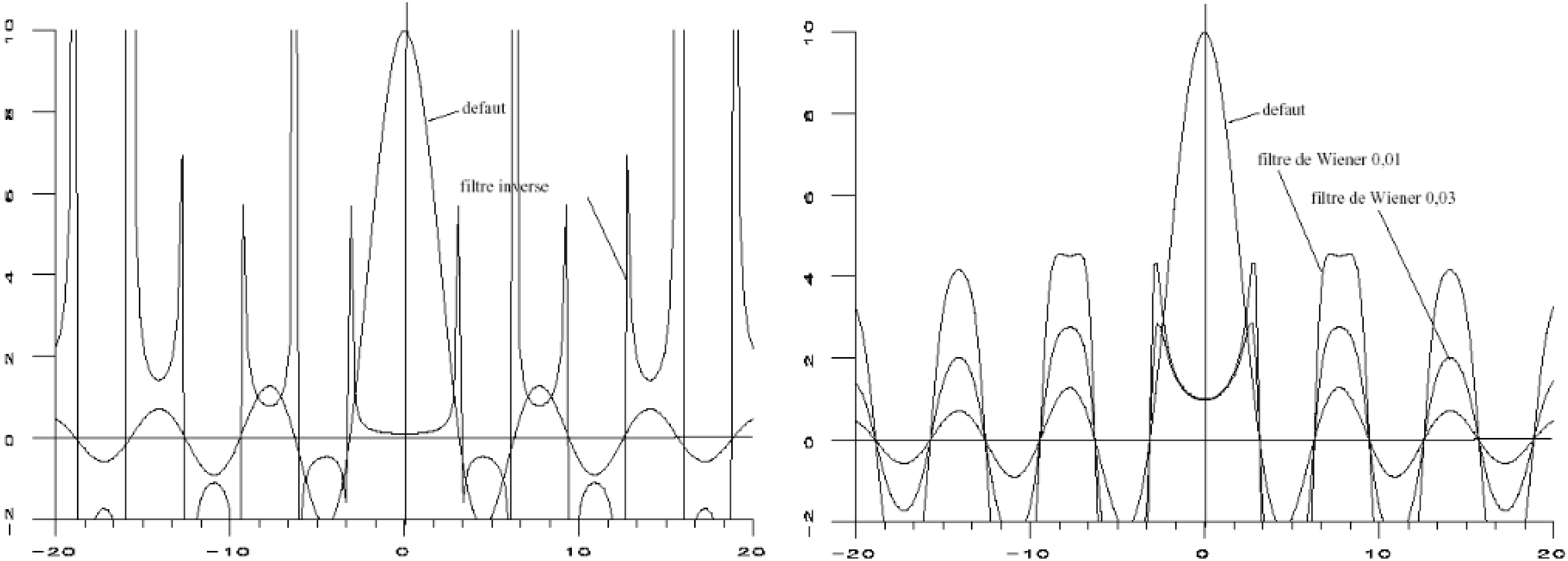
$$F = I \star g_d + b \xrightarrow{\textit{Fourier transform}} \tilde{F} = \tilde{I} \times \widetilde{g_d} + \tilde{b} \xrightarrow{\textit{Wiener filter}} \boxed{\tilde{J_d} = \frac{\widetilde{g_d}' \times \tilde{F}}{\widetilde{g_d}\widetilde{g_d}' + \alpha}} \xrightarrow{\textit{Inverse Fourier transform}} J_d$$

$$\alpha \approx \frac{\left\langle \left|\tilde{b}(u)\right|^2 \right\rangle}{\left\langle \left|\tilde{I}(u)\right|^2 \right\rangle}$$

α is a regularisation term, which depends on the relative power of noise *b* with respect to image signal *I*. It can be set as constant or depend on frequencies: $\alpha(u)$. Wiener filtering thus performs a trade-off between deconvolution and regularisation.

In any case, the reconstruction error ε_d strongly depends on the zeros of the convolution filter in the frequency domain ($\widetilde{g_d}$).

DECONVOLUTION: INVERSE AND WIENER FILTERING



Left: a (constant speed) motion blur in the frequency domain (cardinal sine), and the corresponding inverse filter.
Right: the same default and the correcting Wiener filters for two different values of α assumed constant.

[Figure: Maître 2003]

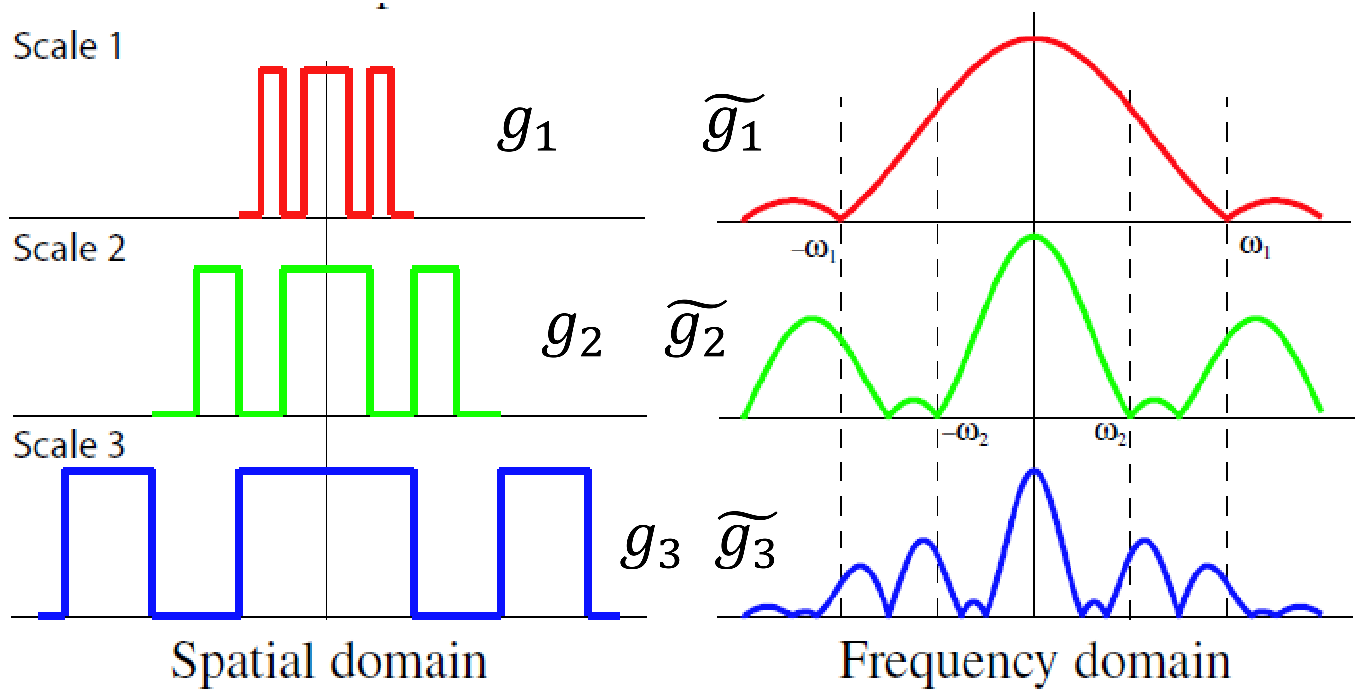
PASSIVE 3D: CODED APERTURE

In deconvolution techniques, the zeros of the filter in the frequency domain are those that mainly contribute to the reconstruction errors.

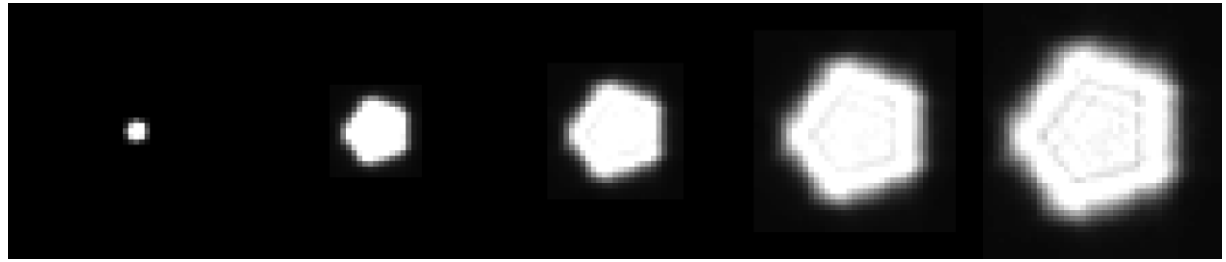
As a consequence, if the different convolution kernel candidates $\{g_d\}_{d \in D}$ have their zeros located at the same frequencies in the Fourier domain, it is much more difficult to distinguish their effects on the image (by deconvolution) than if their zeros appear at different locations.

The principle of coded aperture is to choose the shape of the aperture in such a way that the zeros of the different filters $\{g_d\}_{d \in D}$ appear, depending on *d*, at different location of the frequency domain:

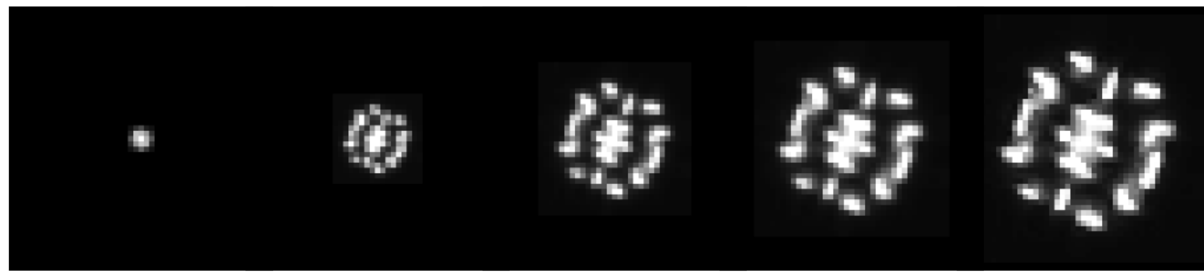
[Levin 2007]



PASSIVE 3D: CODED APERTURE



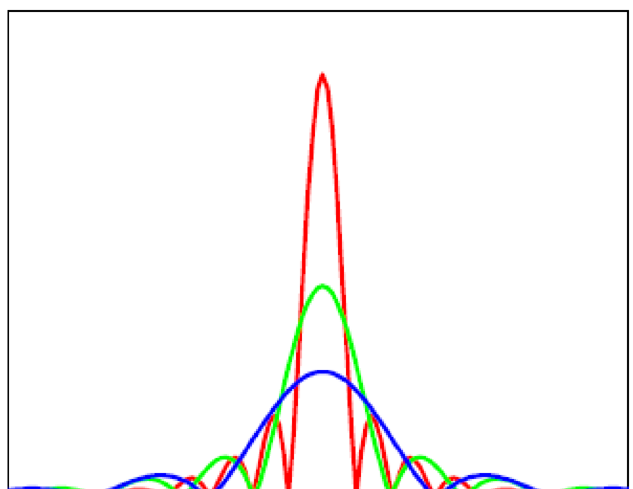
Traditional 5-blade diaphragm and the family $\{g_d\}_{d \in D}$ *of kernels.*



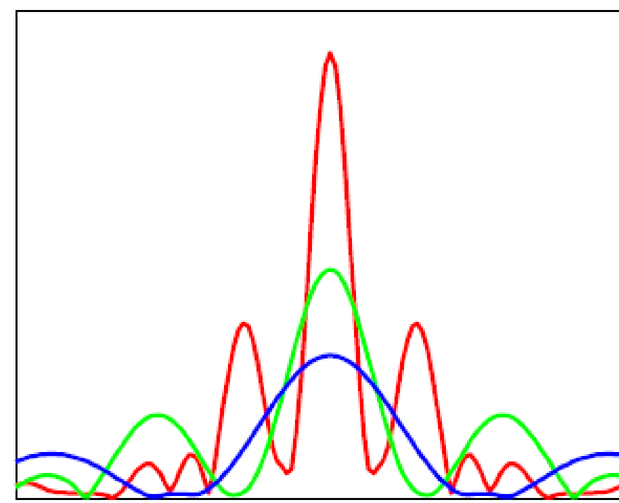
Coded aperture and the family $\{g_d\}_{d \in D}$ *of kernels.*

Comparing the kernels in frequency domain $\{\widetilde{g_d}\}_{d \in D}$ between classic and coded apertures (note the location of the zeros):

[Levin 2007]



Conventional aperture



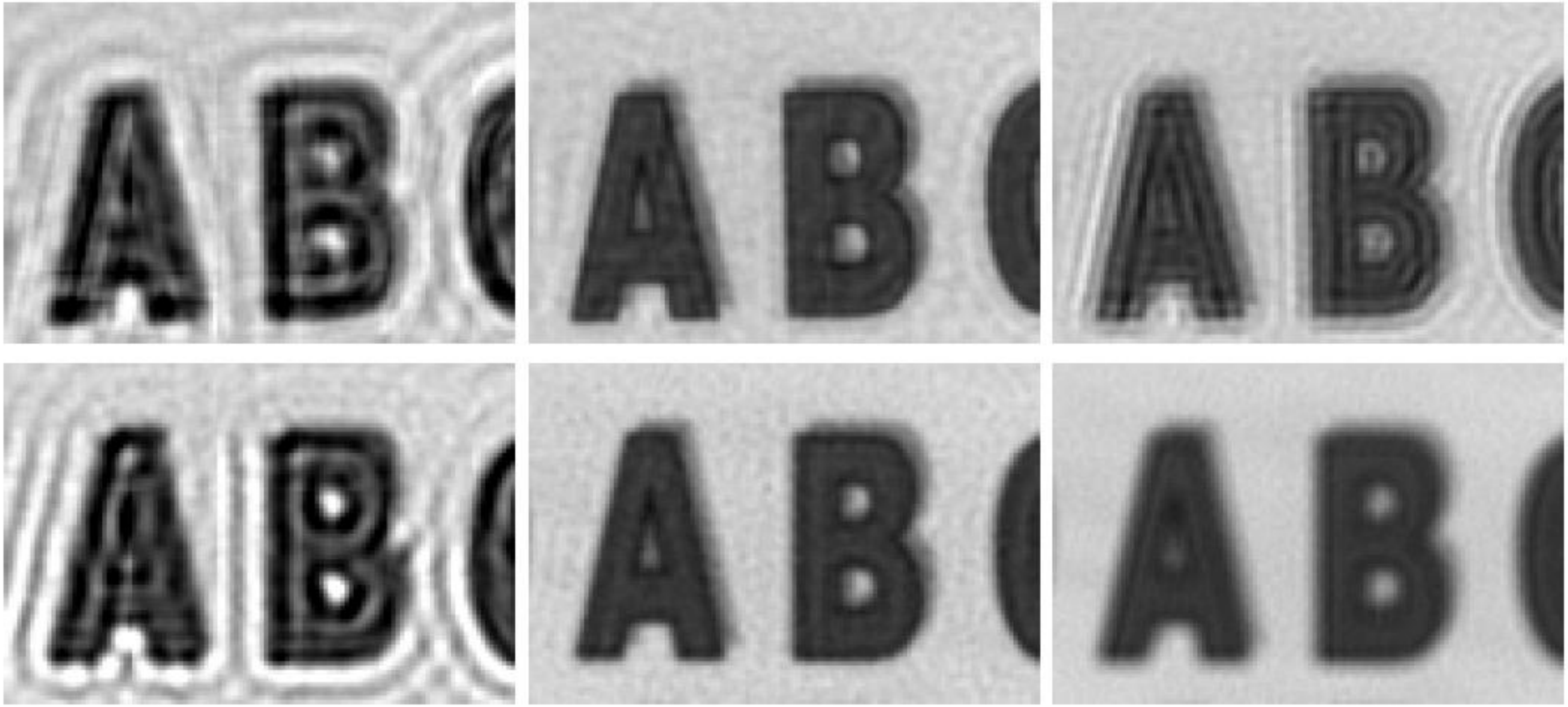
Coded aperture

PASSIVE 3D: CODED APERTURE

Images obtained by deconvolution with coded aperture allow to better discriminate the right scales (distances):

[Levin 2007]

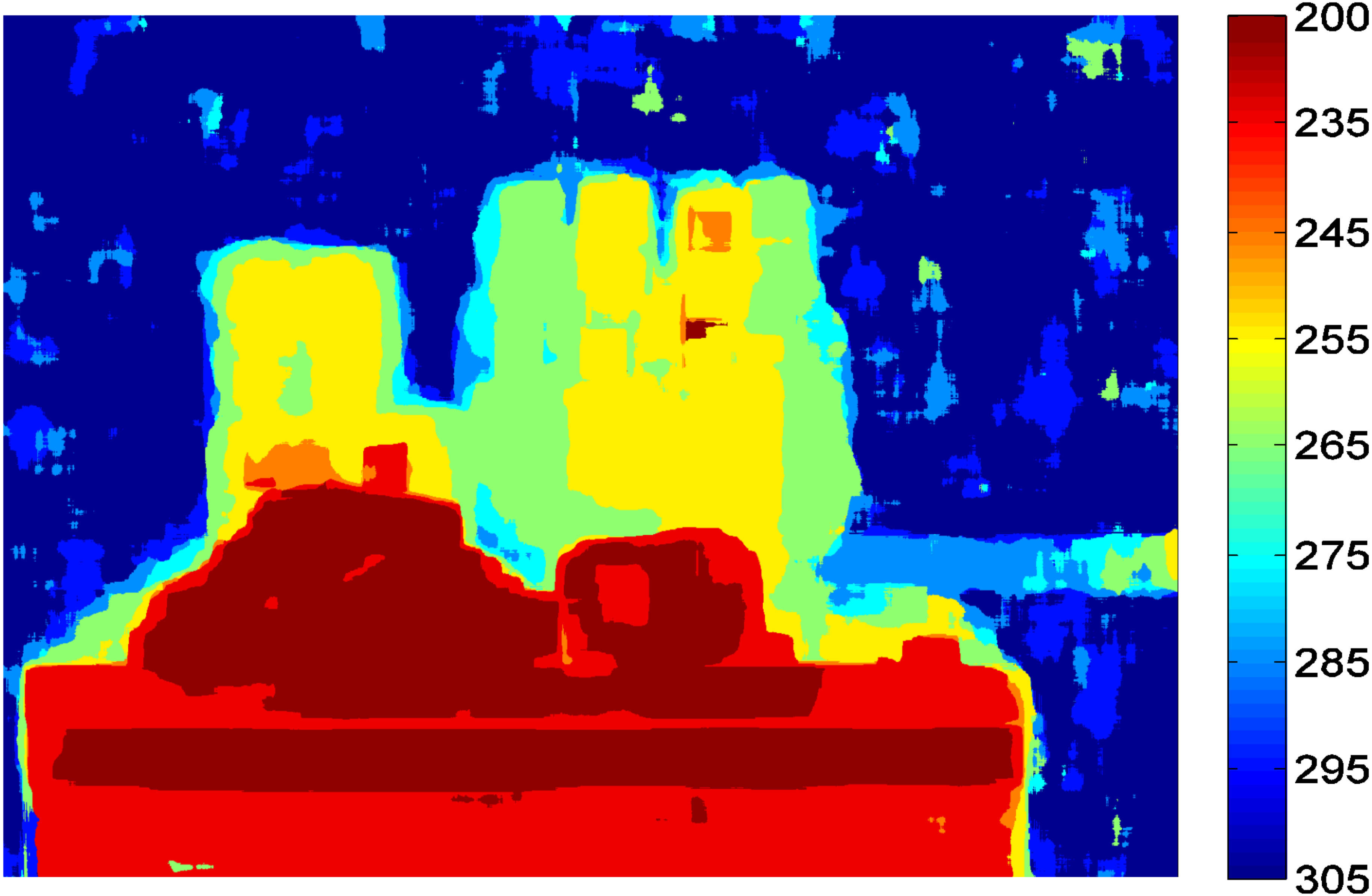
	$d > d_{opt}$	$d \simeq d_{opt}$	$d < d_{opt}$
Coded aperture			
Classic aperture			



RANGE TEST IMAGE

[Levin 2007]

RANGE IMAGE: CODED APERTURE RAW RESULT



[Levin 2007]

RANGE IMAGE: RESULT AFTER POST-PROCESSING

[Levin 2007]

Part 3: ELECTRONIC RETINAS

The goal of co-design techniques is to globally optimise a vision system using an opportunistic approach that makes the most of all different parts of the system and tries to combine them more closely.

Electronic retinas are part of this approach, by extending the electronics of the sensor to the maximum, from image acquisition toward processing, thus performing image analysis within the focal plane.

The interest is to reduce at the minimum the computation time and/or the energy consumption, thanks to:

1. The dramatic reduction of the data flow
2. The use of massive data (i.e. pixel-wise) parallelism

VISION SYSTEMS BOTTLENECK

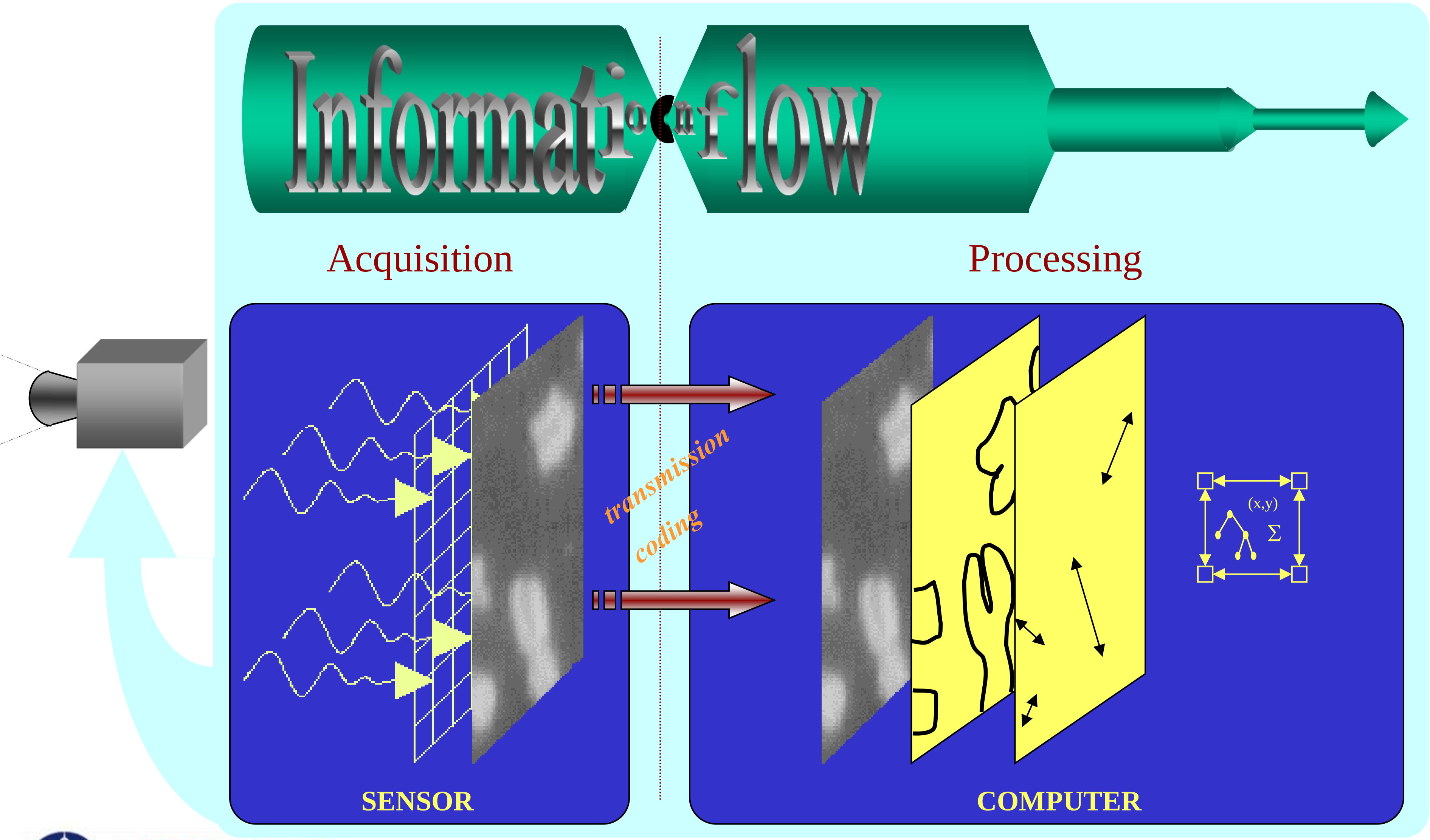
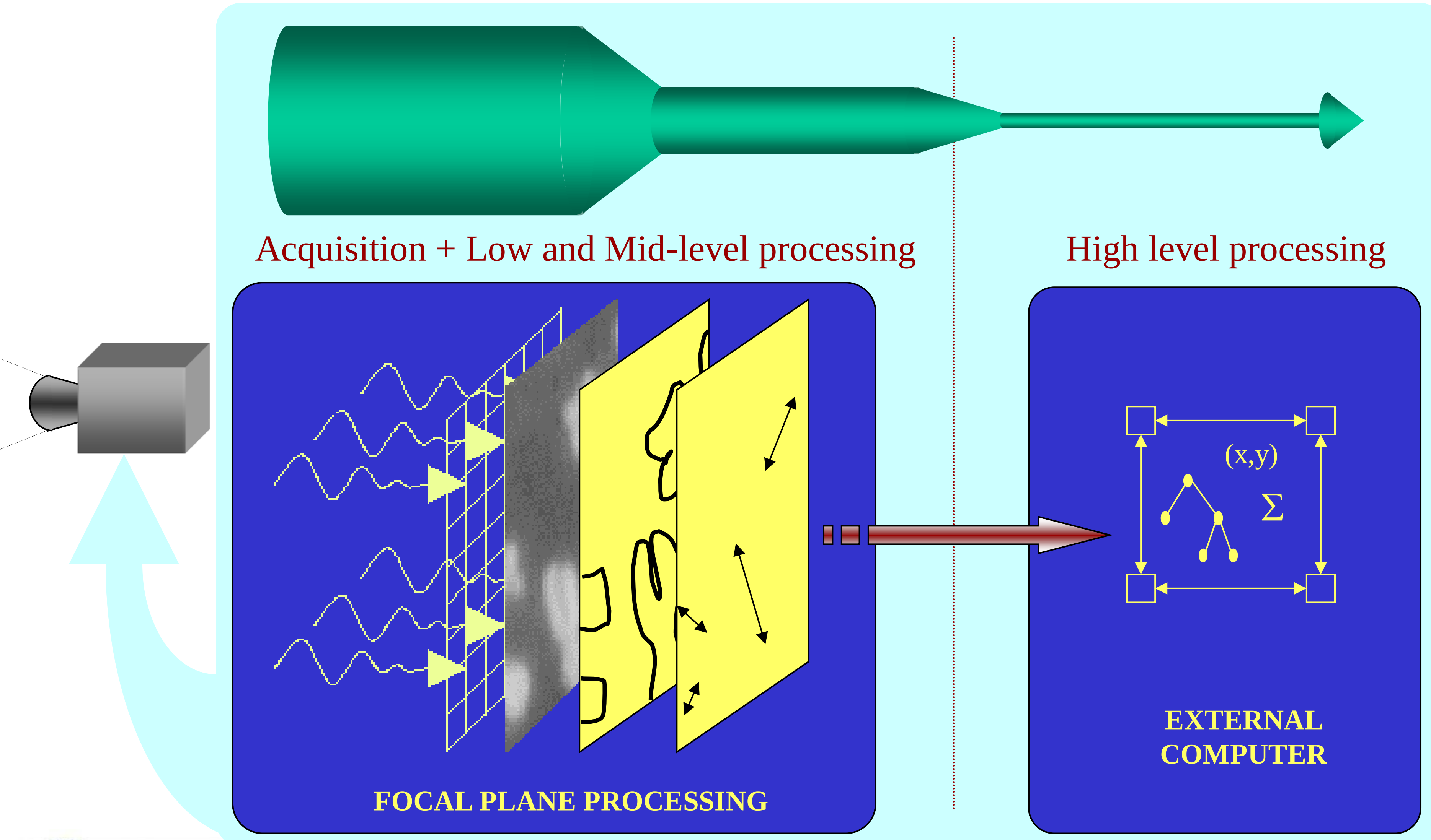
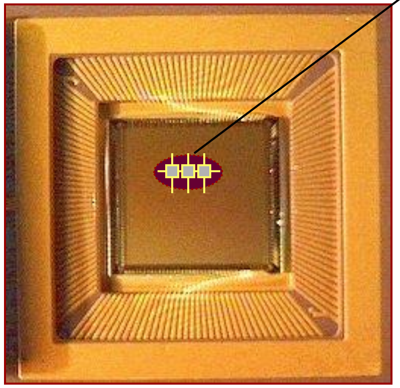
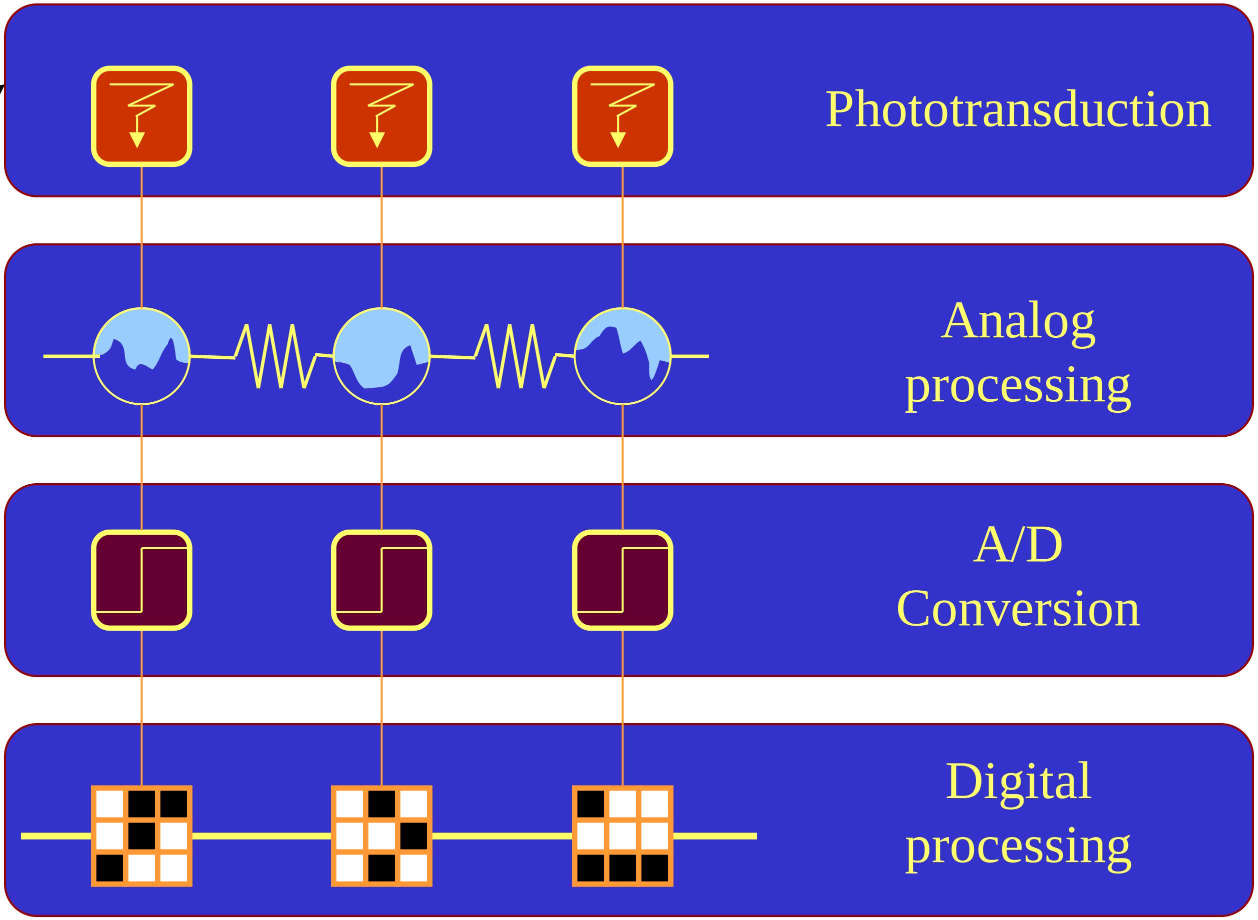


IMAGE PROCESSING WITHIN THE FOCAL PLANE

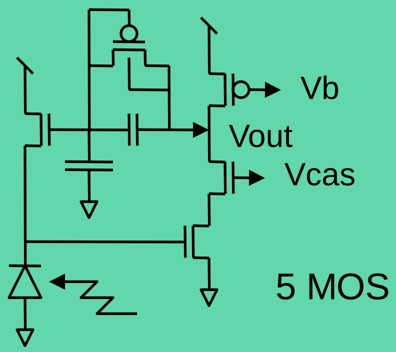


PROGRAMMABLE RETINAS





ANALOG APPROACHES: MOTION DETECTION

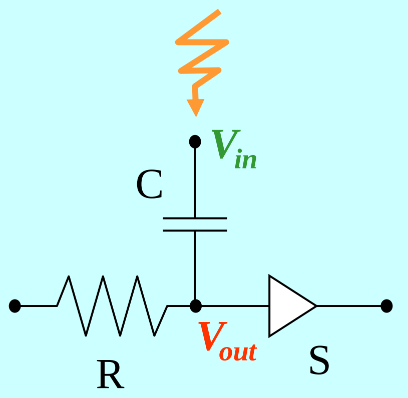


Time change detection [Delbrück 93]

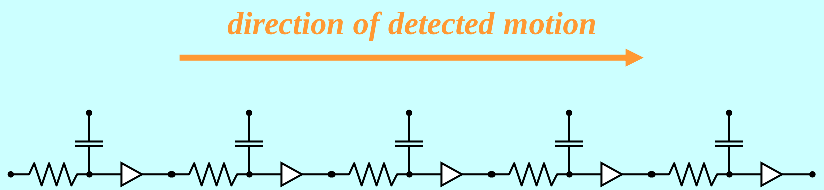
[Figure Th. Bernard 2007]

→ ''Event-based Cameras'', very *low-power*, *ultra-fast* and *asynchronous* ***[Lichtsteiner 2007]***.

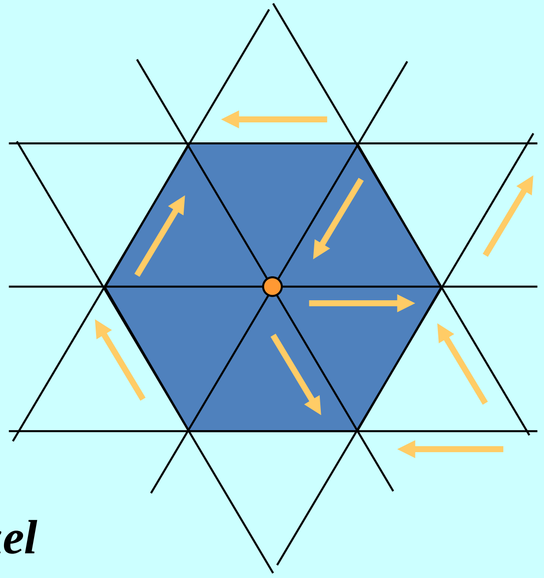
ANALOG APPROACHES: OPTICAL FLOW COMPUTATION



A basic cell



A delay line



A pixel

- The basic cells combines a temporal high-pass filtering of the input signal V_{in} and a low-pass filtering of the output signal from the adjacent cell.
- The delay line then detects a displacement that occurs in the direction of the line.
- At the level of the pixel, the combination of signals measured on each line provides the estimation of the apparent velocity vector (optical flow).

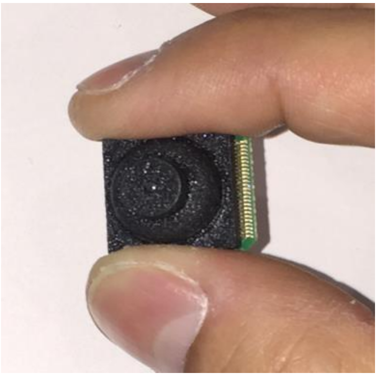
[Delbrück 93]

ANALOG *VS* DIGITAL APPROACHES

ANALOG	DIGITAL
☺ Continuous time, asynchronous	☹ Discrete time, synchronous
☺ Compact solutions for low-level operators, either linear or not, either local or global	☹ Clock sequencing is costly
☺ Very low power	
☹ Dedicated, or hardy configurable devices	☺ Programmable and versatile devices
☹ Reusability (Tool boxes) questionable	☺ Routines, libraries, toolboxes, easy to setup
☹ Scalability w.r.t. Technology difficult	☺ Scalability w.r.t. Technology simpler
☹ Control and Reliability very hard	☺ Control and Reliability measurable

MODERN EVENT-BASED CAMERAS

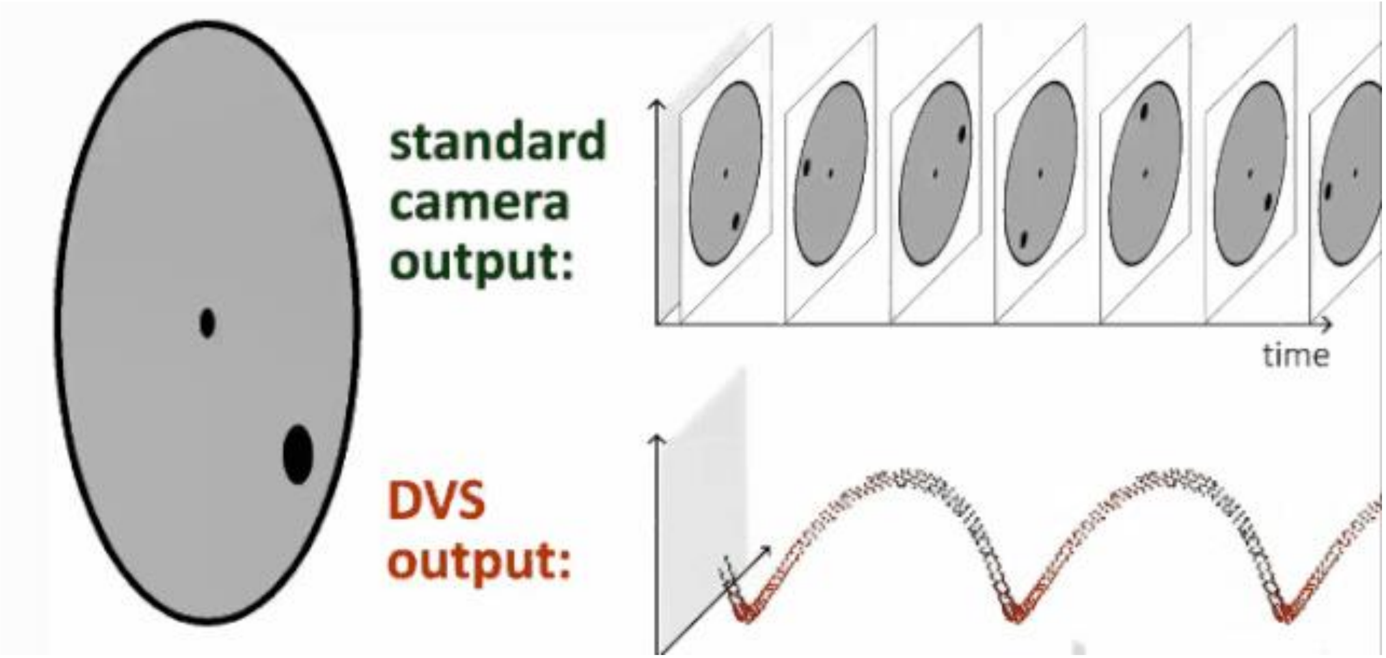
- Novel sensor that measures only **motion in the scene**
- **First commercialized in 2008** by T. Delbruck (UZH&ETH) under the name of Dynamic Vision Sensor (DVS)
- **Low-latency** (~ 1 μs)
- **No motion blur**
- **High dynamic range** (140 dB instead of 60 dB)
- **Ultra-low power** (mean: 1mW vs 1W)



Mini DVS sensor from IniVation.com

[Slide from Scaramuzza's Tutorial 2020]

http://rpg.ifi.uzh.ch/research_dvs.html



SCIENCES SORBONNE UNIVERSITÉ

EVENT-BASED CAMERAS: DYNAMIC RANGE

Low-light Sensitivity (night drive)

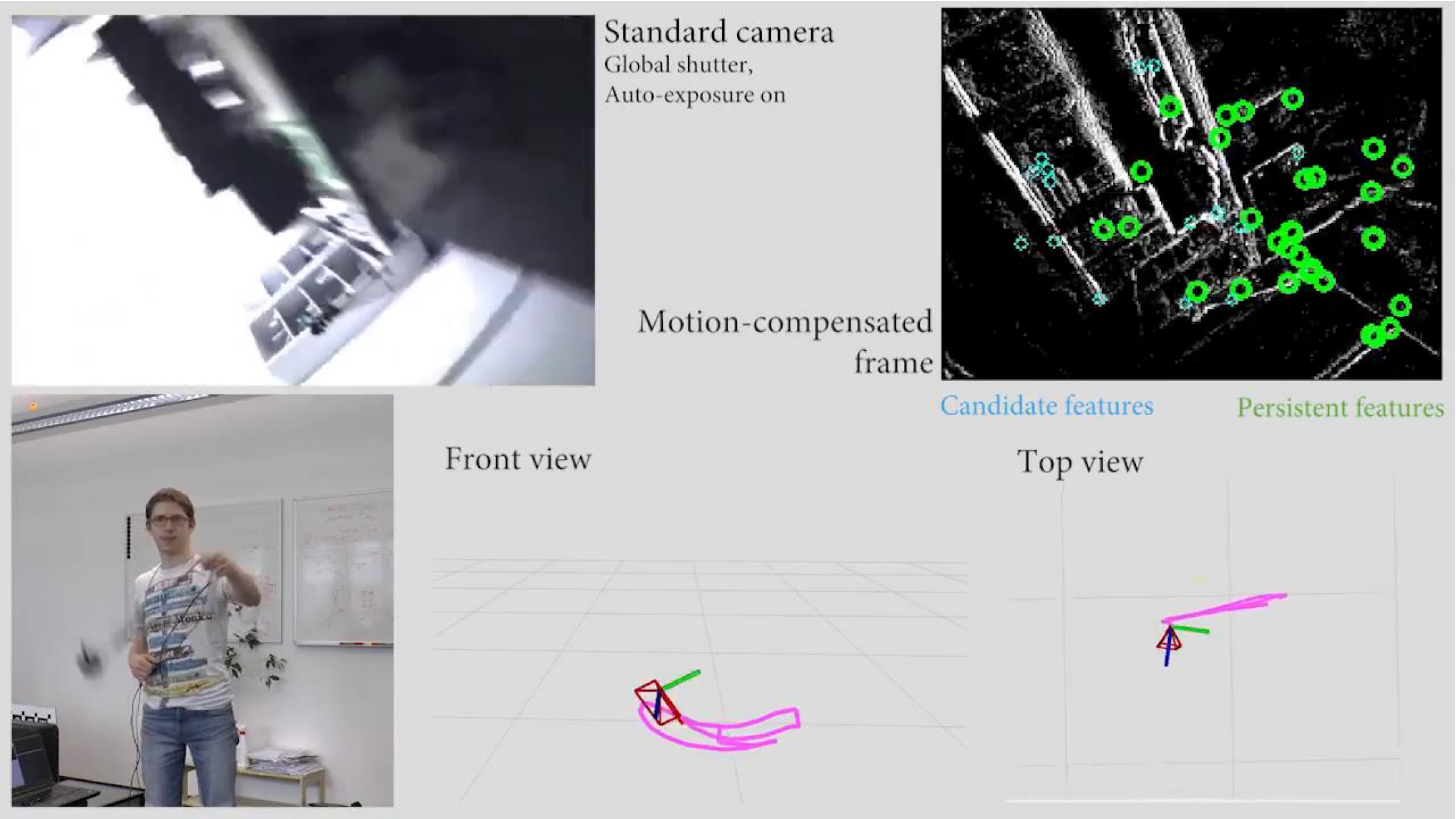
GoPro Hero 6

Event Camera by *Prophesee*
White = Positive events
Black = Negative events

[Slide from Scaramuzza's Tutorial 2020]

http://rpg.ifi.uzh.ch/research_dvs.html

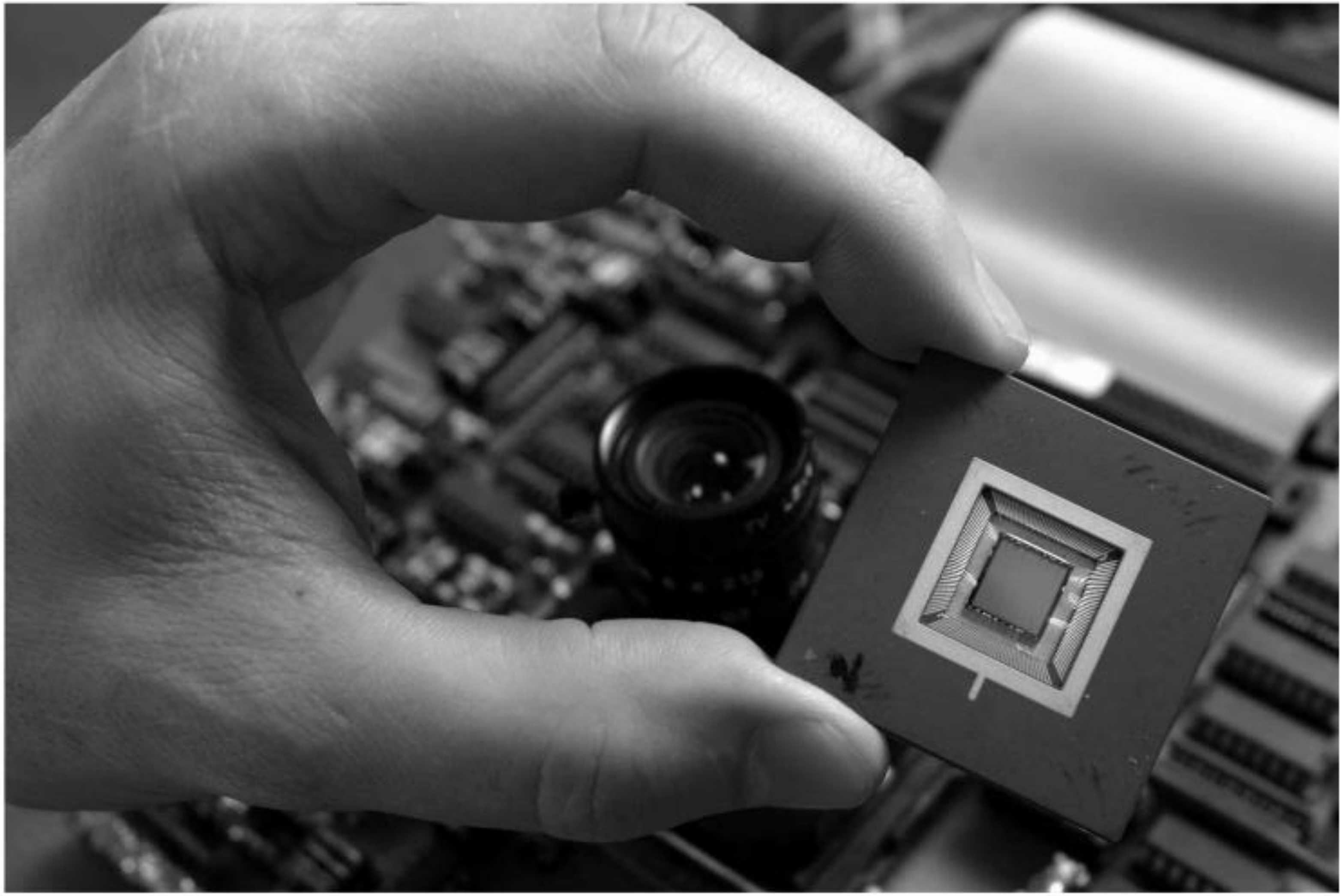
EVENT-BASED CAMERAS: ULTIMATE SLAM



[Slide from Scaramuzza's Tutorial 2020]

http://rpg.ifi.uzh.ch/research_dvs.html

PROGRAMMABLE RETINA *PVLSAR 34* [T. BERNARD 2004]



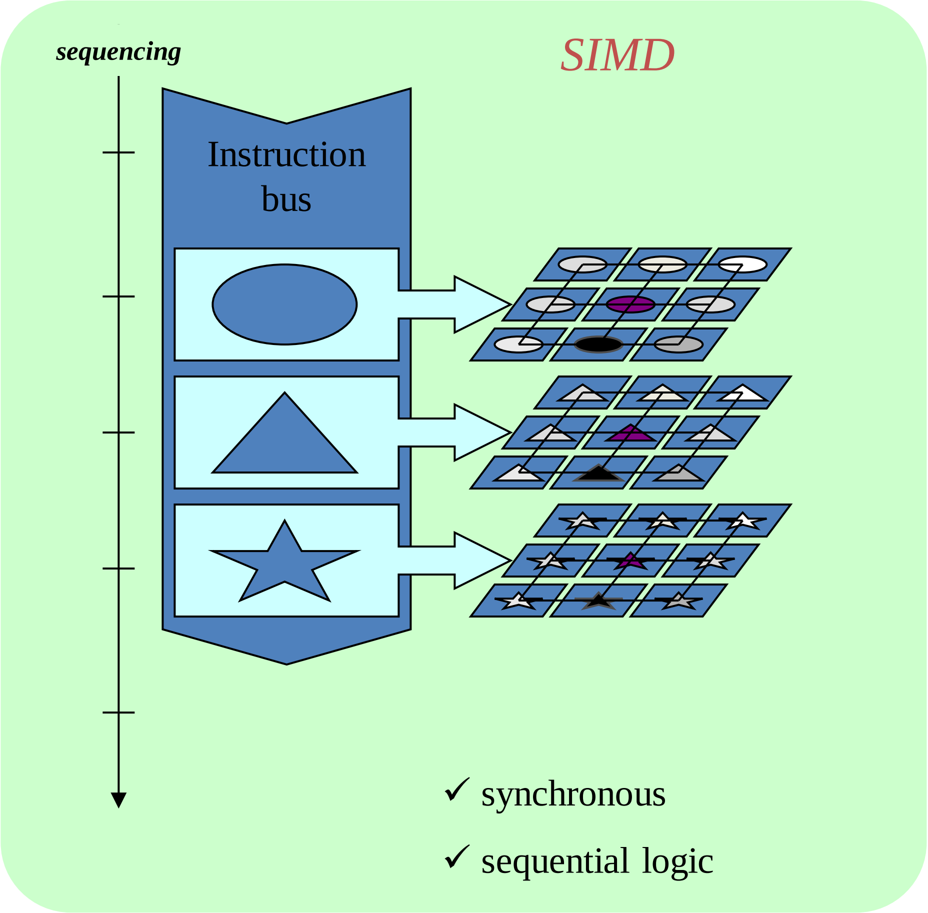
Boolean Retina

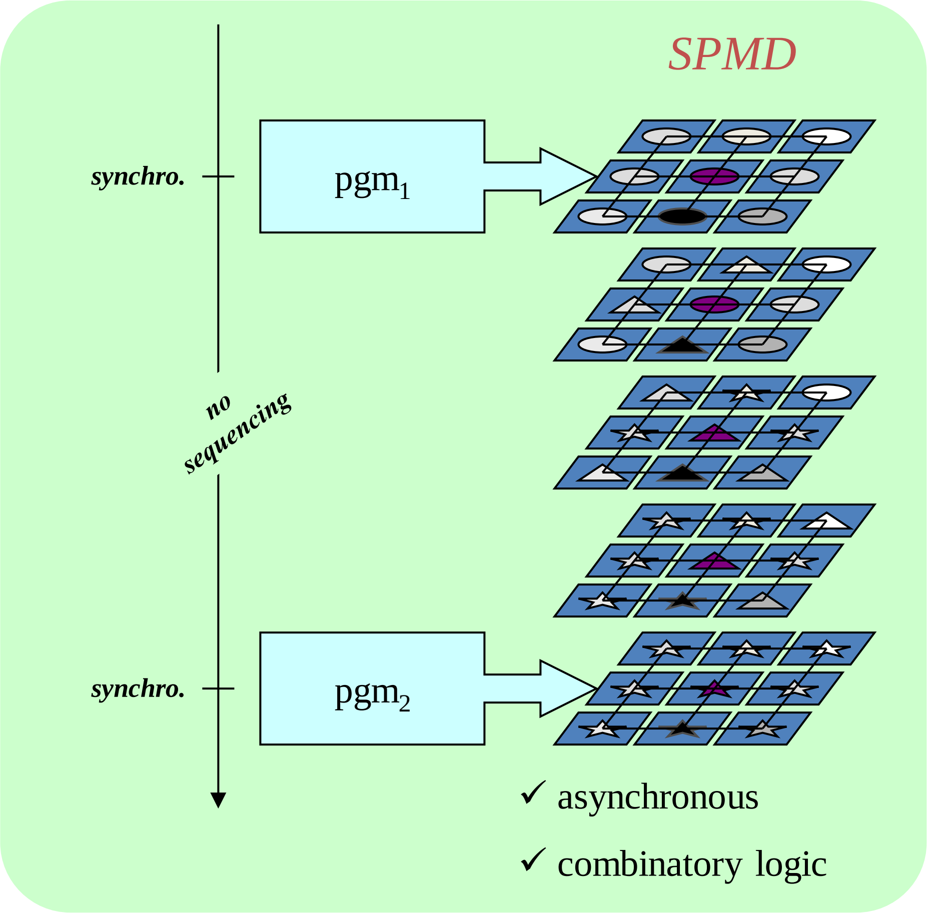
SIMD 200x200

AMS 0,35 µm

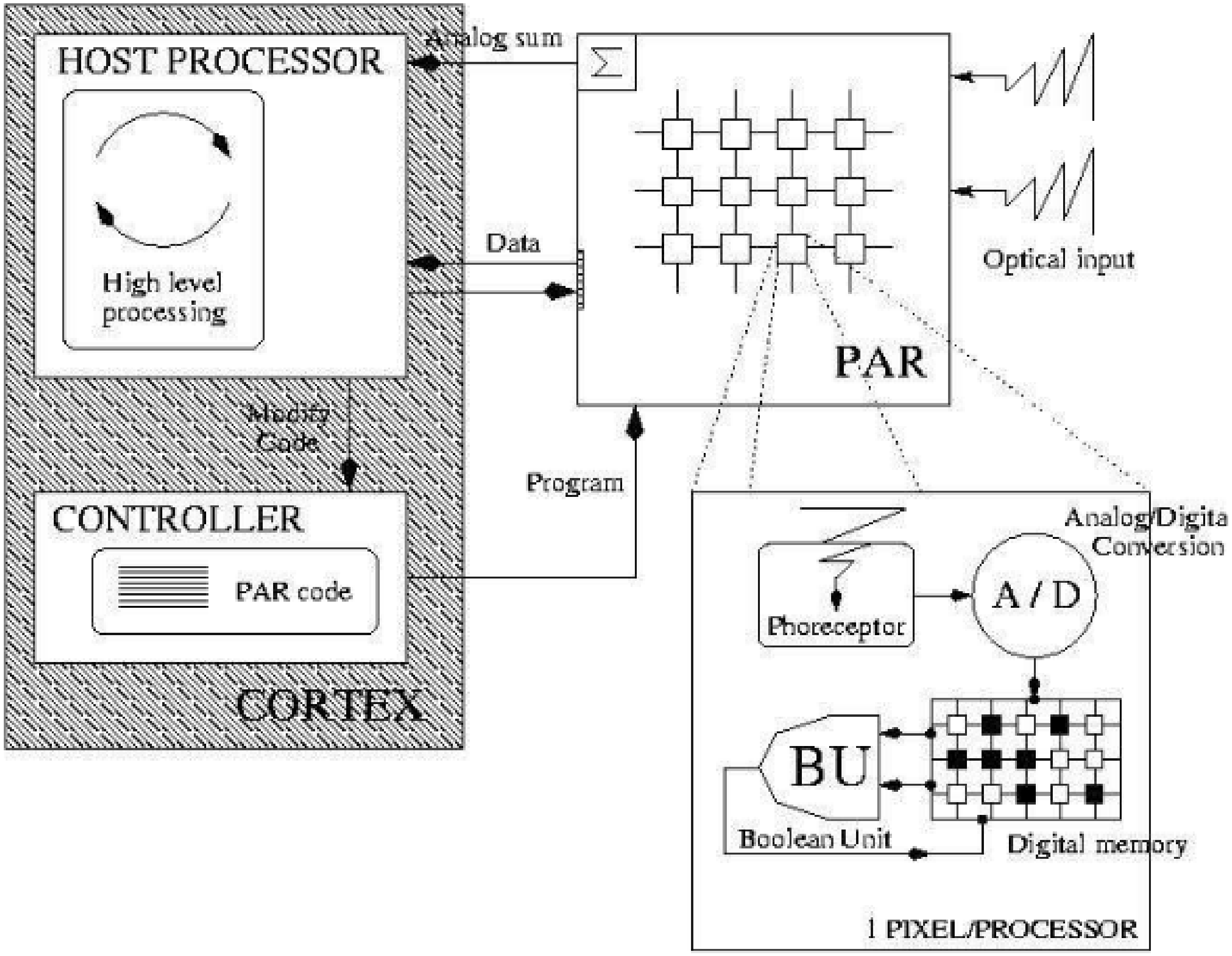
THE PROGRAMMABLE RETINA AS A PARALLEL MACHINE

The working model of the digital retina as a *parallel machine*, is of the *SIMD* (Single Instruction Multiple Data) type, or *SPMD* (Single Program Multiple Data) type.





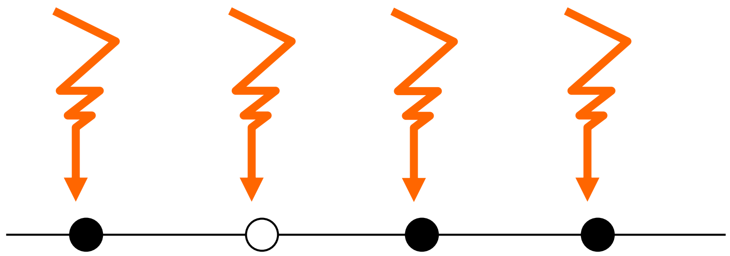
VISION SYSTEM BASED ON ELECTRONIC RETINA



- Heterogeneous architecture
- Hybrid architecture:
  - **Synchronous / Asynchronous**
  - **Digital / Analog**
- Fusion Acquisition / Processing

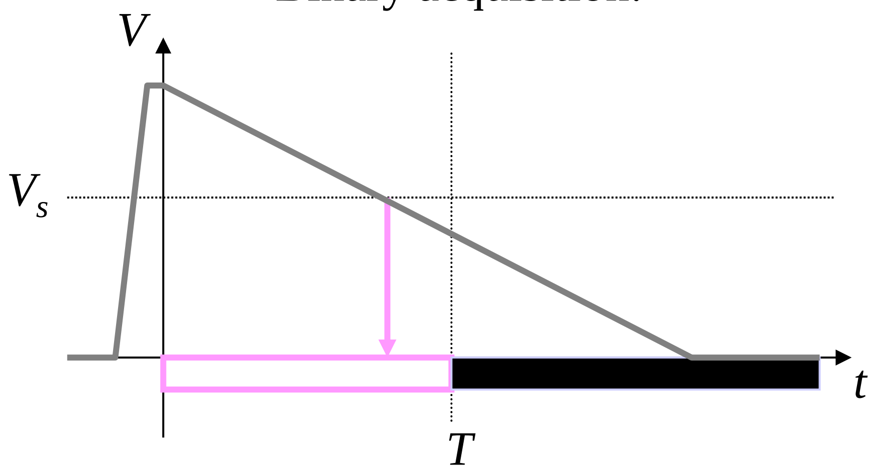
DIGITAL RETINA ORIGIN: LCP RETINA (1993)

1 bit...

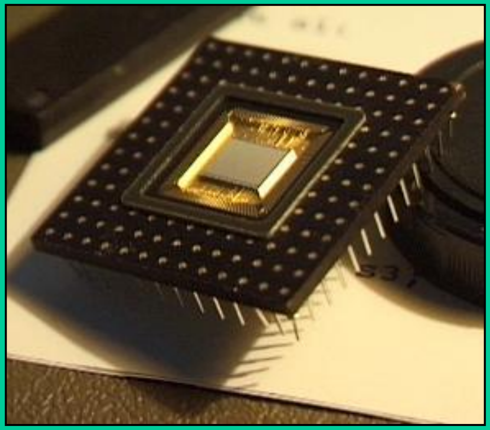


Acquisition plane

Binary acquisition:



Comparing the tension at the bounds of the photodiode at time T to a threshold V_s



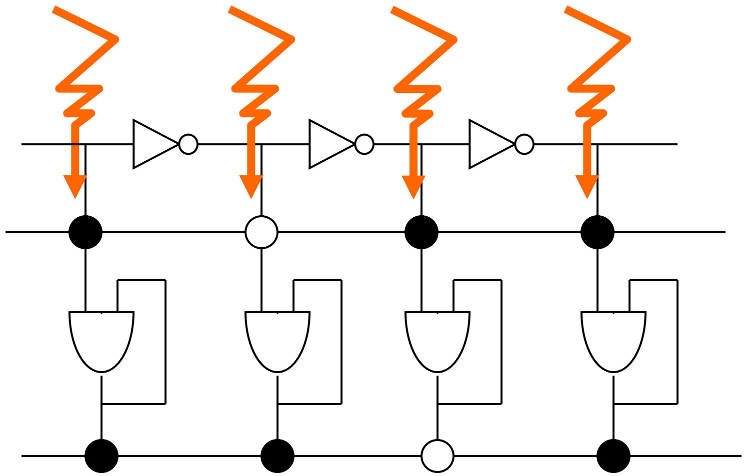
LCP Retina (Bernard - Zavidovique - Devos 1993)

- ✓ 65x76 pixels grid
- ✓ CMOS 2 µm
- ✓ 28 transistors/pixel
- ✓ pixel size: 100x80 μm^2
- ✓ 1 Boolean Unit per pixel
- ✓ memory: 3 bits/pixel

SCIENCES SORBONNE UNIVERSITÉ

DIGITAL RETINA ORIGIN: LCP RETINA (1993)

...2 bits...



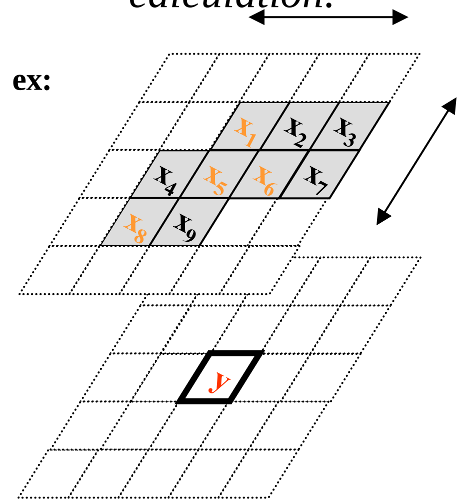
- *Translation (possibly complemented) on plane 1*
- *Computation of the logical AND on plane 2*

Acquisition plane

AND plane

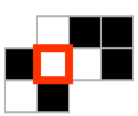
Conjunctive monom calculation:

ex:



$$y = x_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge \bar{x}_4 \wedge x_5 \wedge x_6 \wedge \bar{x}_7 \wedge x_8 \wedge \bar{x}_9$$

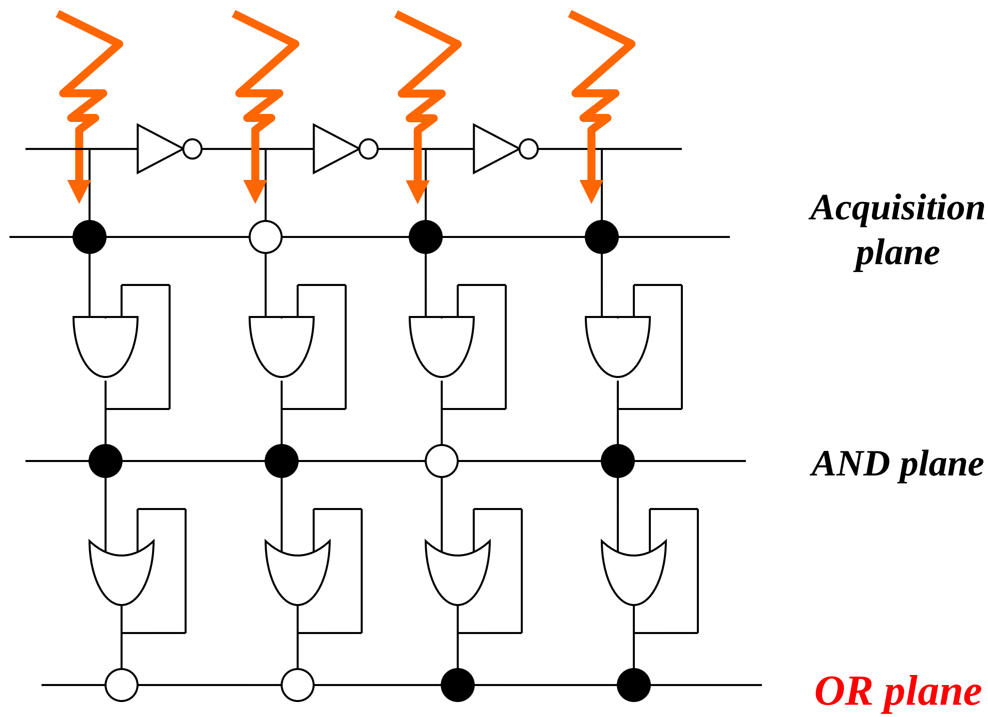
detecting the presence of a certain pattern within the binary image:



Hit-or-Miss Transform!

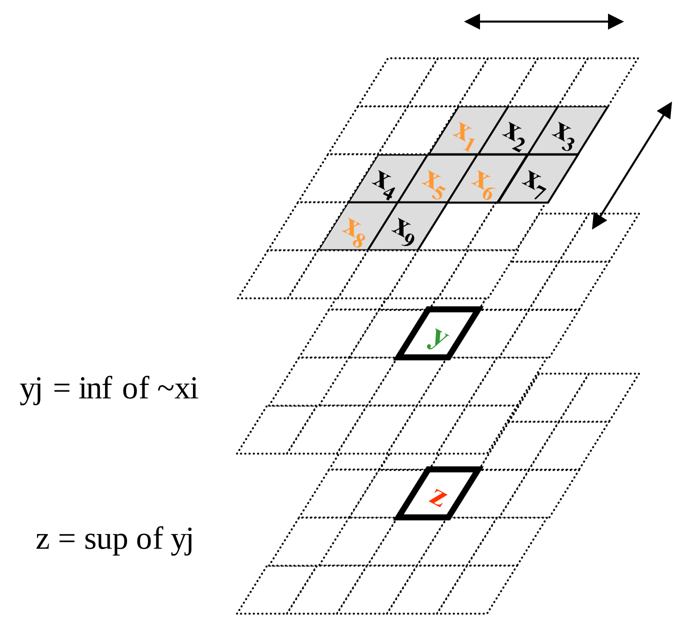
DIGITAL RETINA ORIGIN: LCP RETINA (1993)

...3 bits!



- *Computation of the logical OR on plane 3*

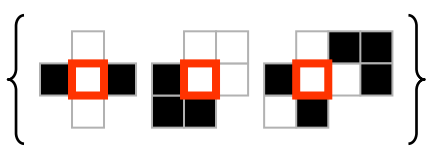
Disjunction of conjunctive monoms:



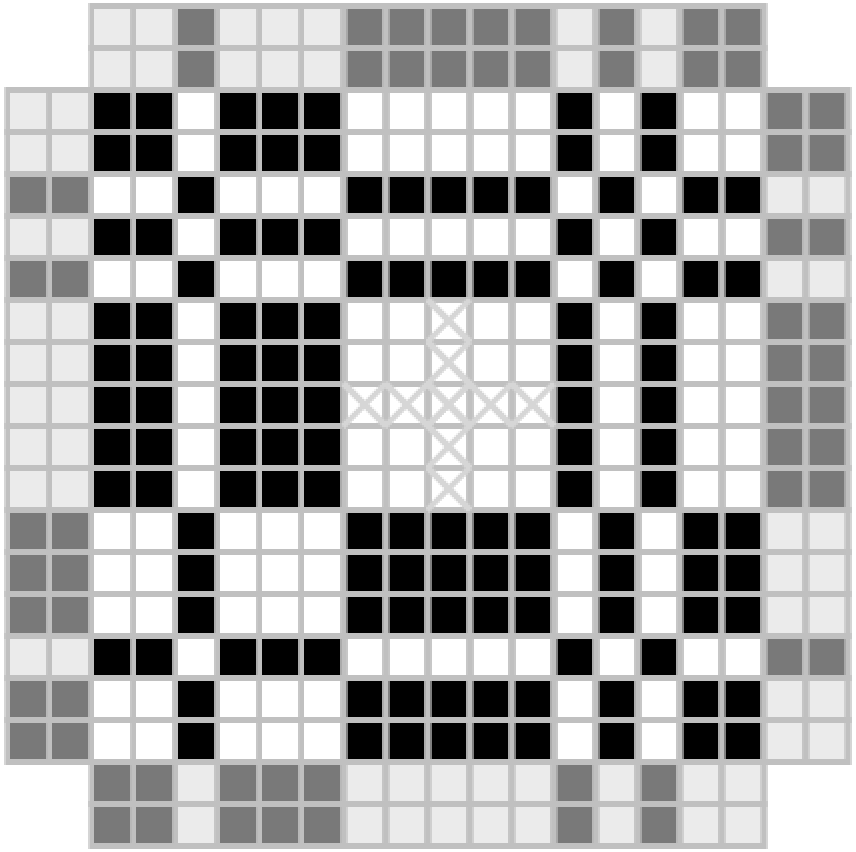
Disjunctive Form

Universal Boolean machine!

detecting the presence of one these patterns within the binary image:



PIXEL POSITION ENCODING FOR BOOLEAN RETINAS



[Bernard 1996]

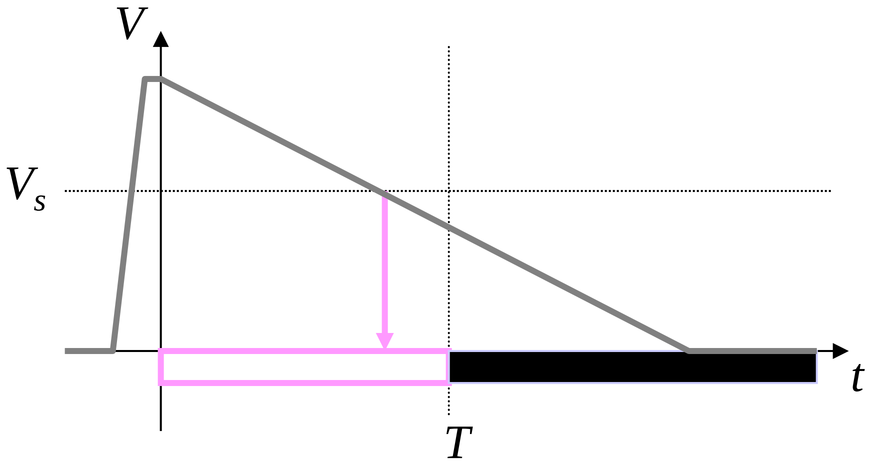
Using De Bruijn 2d sequences, a digital retina only needs one bit of memory per pixel to locally encode the position of each pixel.

Figure: *B*(2,9), using a cross-shaped neighbourhood.

SCIENCES SORBONNE UNIVERSITÉ

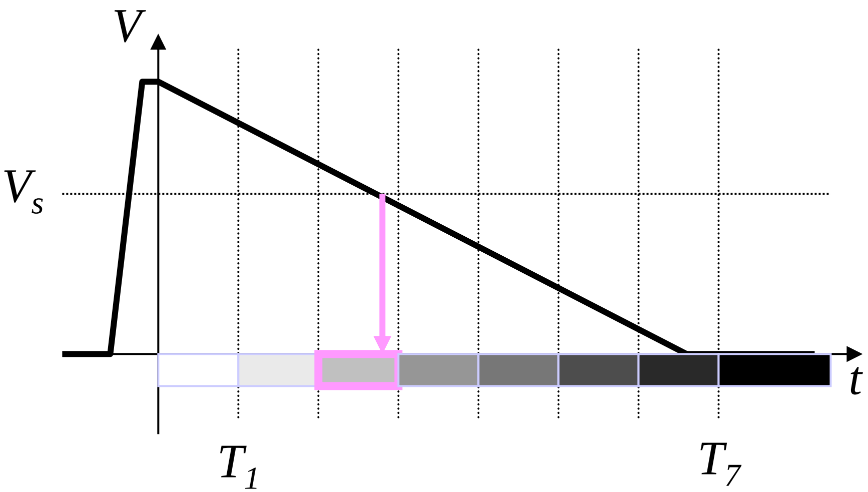
ACQUISITION AND ANALOG/DIGITAL CONVERSION

Binary acquisition by thresholding:



Comparing the tension at the bounds of the photodiode at time T to a threshold V_s

Gray level acquisition by multiple thresholding:

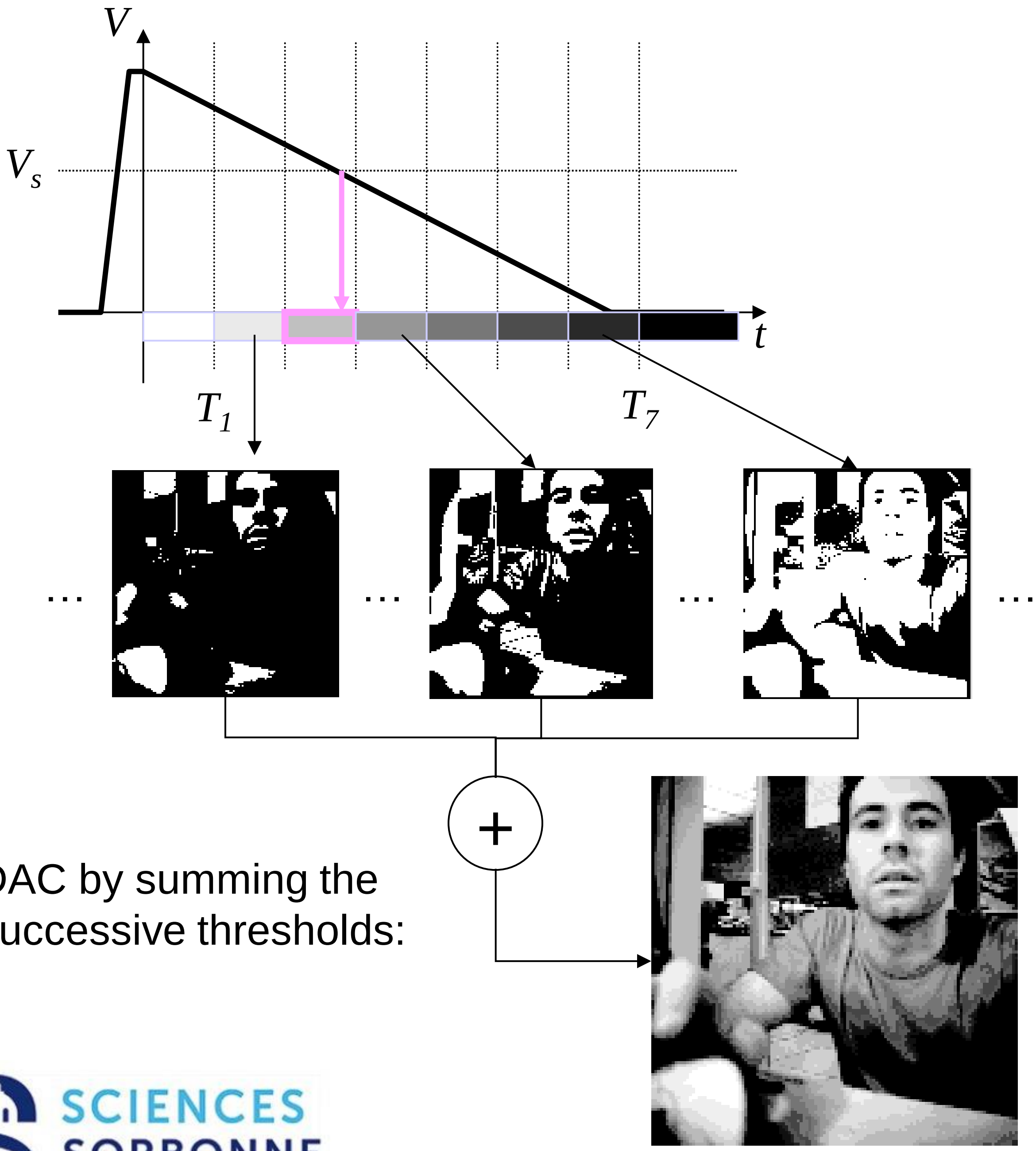


Comparing the tension at the bounds of the photodiode at n times T_i to the threshold V_s

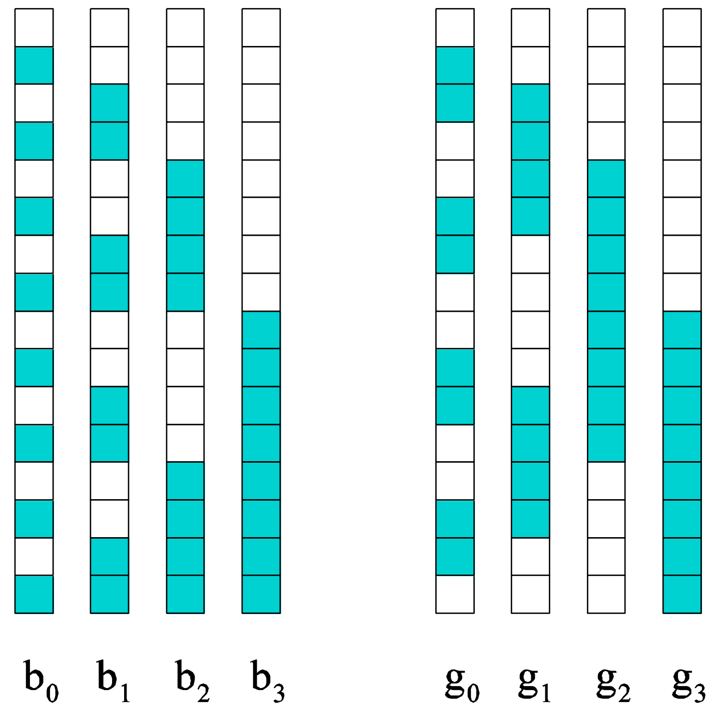
NSIP Process
(Near Sensor Image Processing):

(Eklund - Svensson - Aström 1996)

ACQUISITION AND ANALOG/DIGITAL CONVERSION



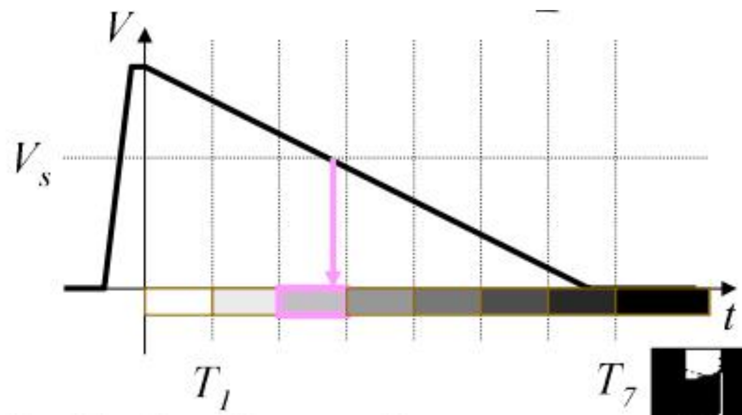
DAC by summing the successive thresholds:



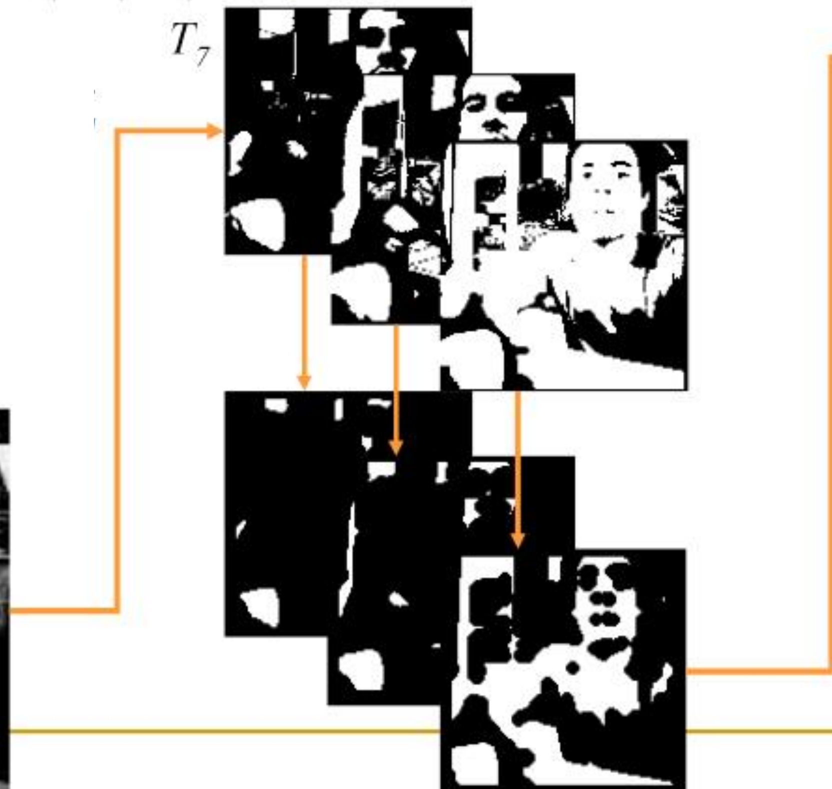
Natural code: log2(n) operations per threshold

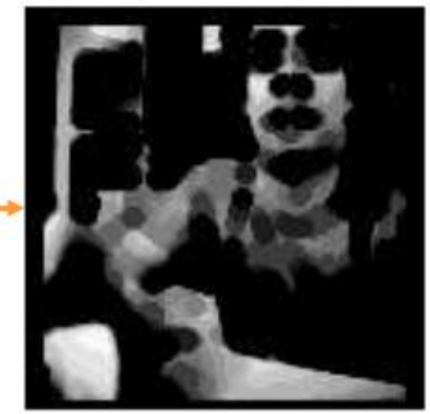
Gray code: one single operation per threshold

FUSION ACQUISITION/PROCESSING



Graylevel acquisition by multiple measure of the photodiode tension along time.





Active sensing:

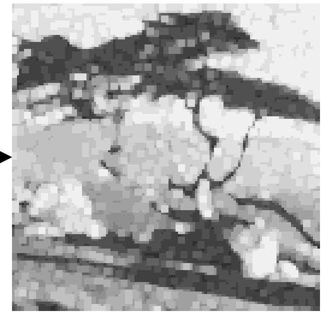
- ***Adaptation to light***
- ***Logarithm compression***
- ***Gain control***

SCIENCES SORBONNE UNIVERSITÉ

EX #1: FROM BOOLEAN COUNT TO RANK FILTERS

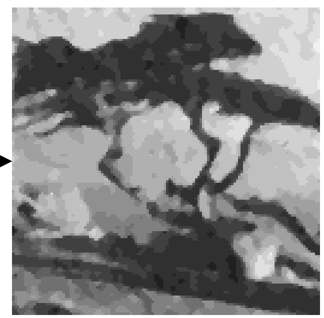
Noisy image I

T_9^1



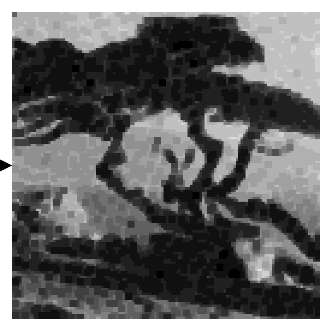
morphologic dilation δ(I)

T_9^5



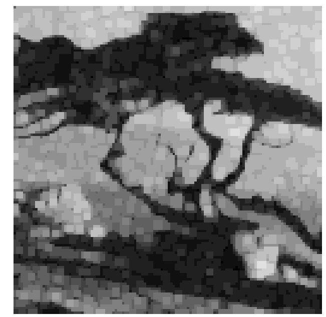
médian filter μ(I)

T_9^9

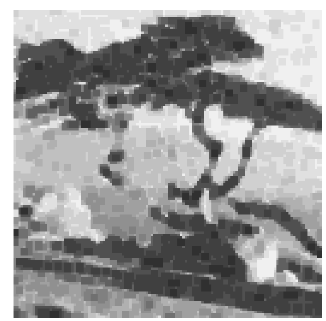


morphologic erosion ε(I)

and by composition/difference:

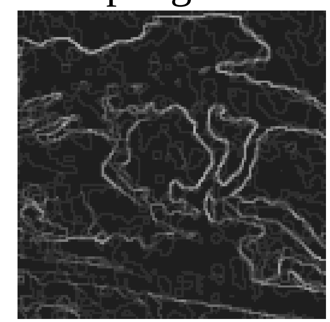


opening δ(ε(I))



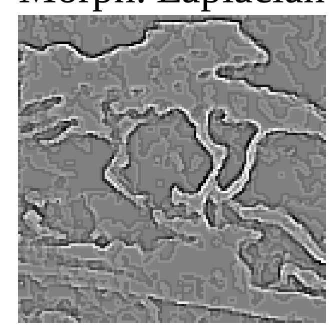
closing ε(δ(I))

Morph. gradient



μ(I)−ε(μ(I))

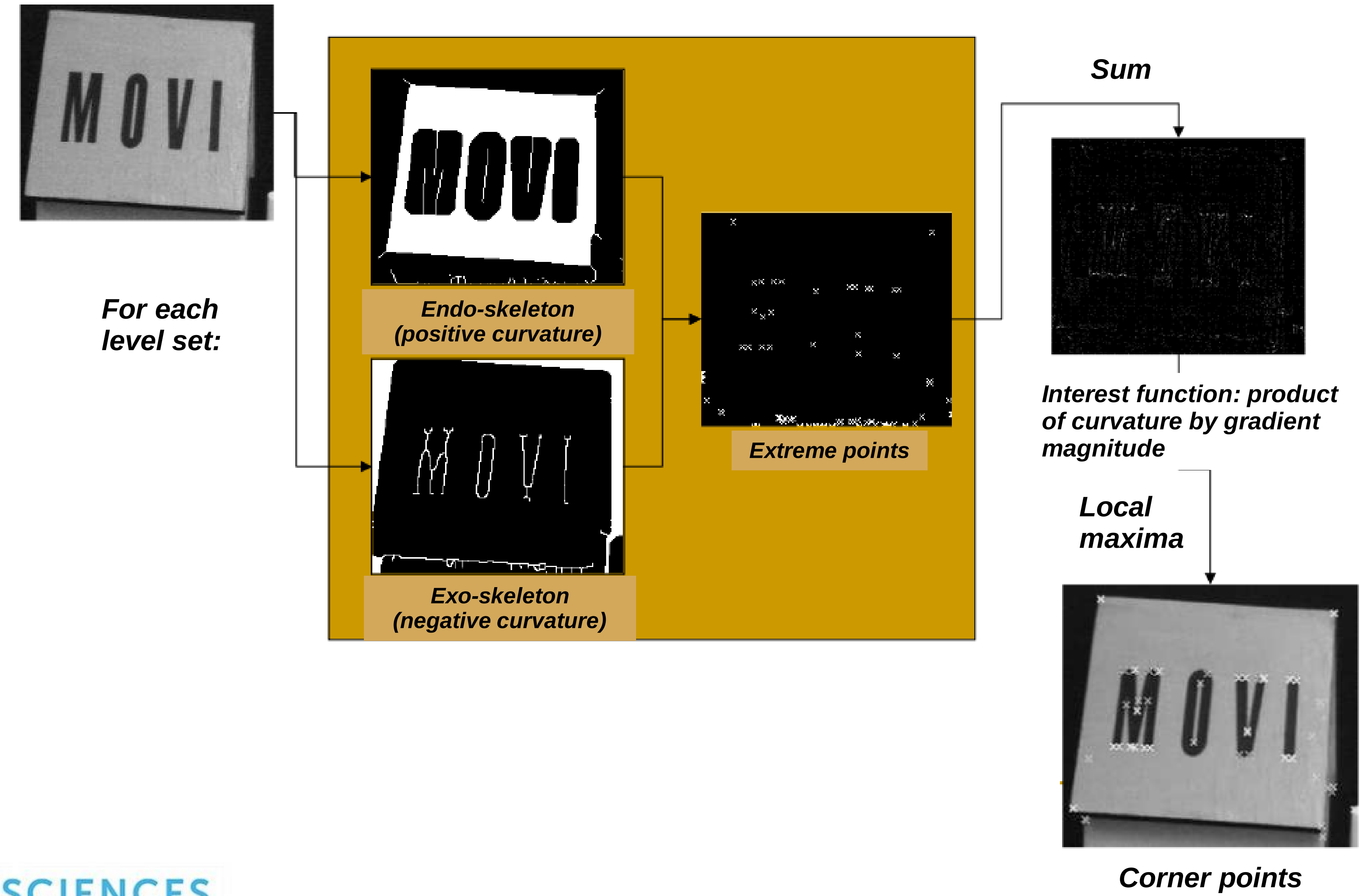
Morph. Laplacian



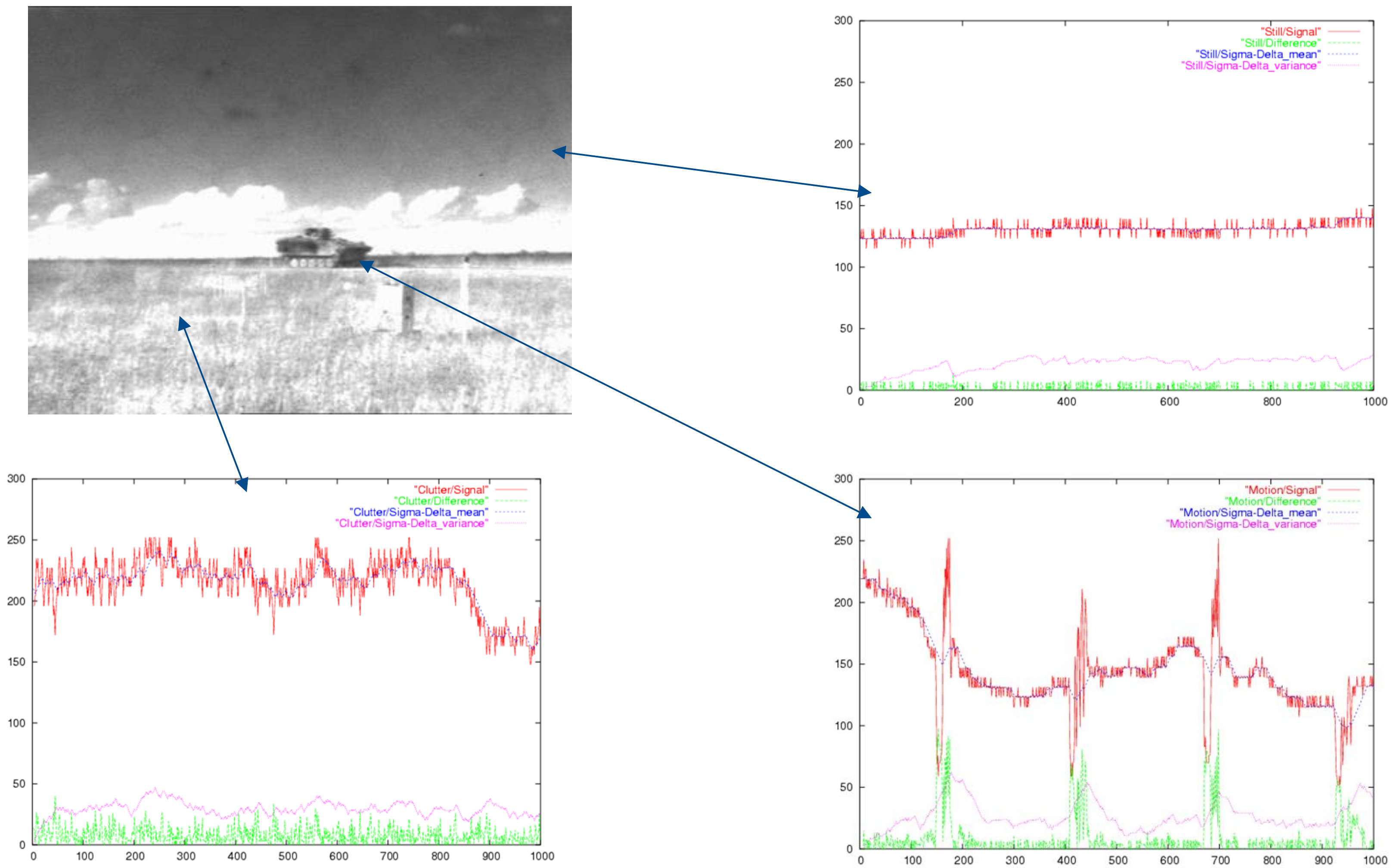
ε(μ(I))+δ(μ(I))-2μ(I)

SCIENCES SORBONNE UNIVERSITÉ

EX #2: MORPHOLOGIC CORNER POINTS



MOTION DETECTION ON *PVLSAR 34…*



…See lecture on *Motion Detection*

SCIENCES SORBONNE UNIVERSITÉ

COURSE #2: CONCLUSIONS AND TAKE-AWAY

Co-design techniques aim at globally optimising a vision system through an opportunistic approach that makes the most of all elements and tries to combine them as closely as possible: optic, mechanics, photo-sensing, digitalisation, processing…

This lecture focused on 3d vision and motion analysis. Many other examples of camera co-design can be found in the domain on computational photography, to improve and “augment” digital images.

In every co-designed system, there exist a balance between, on the one hand the hardware complexity and the active/intrusive character of the system (lighting…), and on the other hand the software complexity.

Nonetheless, the weight of the software remains important in most presented systems.

Another important point to keep in mind: the difficulty (if not impossibility) for passive systems, to deal with untextured areas (homogeneous zones).

The underlying scientific principles of the presented systems are generally quite old, but their technological maturity is recent, and new off-the-shelf products are released regularly…

REFERENCES (Part 1)

[Chiabrando 2009] F. Chiabrando, R. Chiabrando, D. Piatti, F. Rinaudo, *Sensors for 3D Imaging: Metric Evaluation and Calibration of a CCD/CMOS Time-of-Flight Camera*, Sensors, vol. *9*, 10080-10096, 2009.

[Geng 2011] Jason Geng, *Structured-light 3D surface imaging: a tutorial*, Advances in Optics and Photonics, vol. *3*, 128-160, 2011.

[Posdamer 1982] J. L. Posdamer and M. D. Altschuler, *Surface measurement by space-encoded projected beam systems*, Comput. Graph. Image Processing 18, (1), 1–17 1982.

[Narasimhan 2006] S. Narasimhan, *Computer Vision: Spring 2006, lecture n.17*, Carnegie Mellon University.

[Zhang 2002] L. Zhang, B. Curless, S. M. Seitz, *Rapid shape acquisition using color structured light and multi-pass dynamic programming*, IEEE Int. Symp. on 3D Data Processing Visualization and Transmission, pp. 24–36, 2002.

REFERENCES (Part 2)

[Adelson 1992] E. H. Adelson, J. Y. A. Wang, *Single Lens Stereo with a Plenoptic Camera*, IEEE Trans. Pattern Analysis and Machine Intelligence 14(2): 99-106, 1992.

[Ng 2005] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan. *Light Field Photography with a Hand-Held Plenoptic Camera,* Stanford University Computer Science Tech Report CSTR 2005-02, April 2005.

[Pentland 1987] Alex P. Pentland, *A new sense for depth of field*, IEEE Trans. Pattern Analysis and Machine Intelligence 9(4): 523-531, 1987.

[Maître 2003] Henri Maître (ss la direction de), *Le Traitement des Images,* Chapitre 5 : Restauration, Hermès – Lavoisier, Série I2C, 2003.

[Levin 2007] A. Levin, R. Fergus, F. Durand, W.T. Freeman, *Image and depth from a conventional camera with a coded aperture,* ACM Transactions on Graphics 26 (3): 70-78, 2007.

SCIENCES SORBONNE UNIVERSITÉ

REFERENCES (Part 3)

[Delbruck 1993] Toby Delbrück, *Silicon retina with Correlation-Based, Velocity-Tuned Pixels*, IEEE Transactions on Neural Networks, Vol. 4, No. 3, pp. 529–541, 1993.

[Lichtsteiner 2007] P. Lichtsteiner, C. Posch, and T. Delbruck. *A 128x128 120dB 15us latency asynchronous temporal contrast vision sensor*. 43. 566-576, 2007.

[Gallego 2020] G. Gallego et al, *Event-based vision: a survey* , IEEE Transaction on Pattern Analysis and Machine Intelligence, Jul. 2020.

[Bernard 1993] T.M. Bernard, B. Zavidovique, and F.J. Devos, *A programmable artificial retina*, IEEE Journal of Solid-State Circuits, 28(7), Jul 1993, p. 789-798.

[Bernard 1996] T.M. Bernard and J.C. Meier, *Cursor-Injective Two-Valued Lattices for a Local Encoding of Pixel Position*, Proc. SPIE, Vol. 2950, Advanced Focal Plane Arrays and Electronic Cameras, 230-241,1996.

[Astrom 1996] A. Aström, R. Forchheimer, and J.E. Eklund, *Global feature extraction operations for near-sensor image processing*, IEEE Transactions on Image Processing, 5(1), 102-110, 1996.

[Lacassagne 2009] L. Lacassagne, A. Manzanera, J. Denoulet, and A. Mérigot, *High performance motion detection: some trends toward new embedded architectures for vision systems.* Journal of Real Time Image Processing, 4(2), 2009, pp. 127--146.

SOME SUGGESTIONS FOR ORAL PRESENTATIONS…

- Vision systems inspired by **bees or flies**, e.g. Biorobotics team from the Institut des sciences du mouvement (Jules Marey), in Marseille…
- **Time-before-contact**: dedicated implementations, or biological studies…
- **Movement and Gestalt**: other examples of perceptual grouping or simplification, or relations with other aspects of Gestalt…
- **Accomodation /Autofocus**: biological mechanisms, opto-mechanics, algorithms…
- **Random dot (auto)stereograms**: creation, matching, relation with textures…
- **Perspective and Texture Gradients** : methods inspired by descriptive geometry for drawing, 3d reconstruction from the perspective…
- Active vision systems using exploration mechanisms inspired by human **ocular movements…**
- **Super-resolution** systems based on micro-movements (micro-saccades)…
- **Saccadic masking**: its use in an active vision system…

SOME SUGGESTIONS FOR ORAL PRESENTATIONS…

- **Shutter based dephasing** measures for a Time-of-Flight camera, e.g. Kinect v2…
- **Using cast shadows or natural lighting** as structured light for 3d reconstruction…
- Structured light: properties of **2d pseudo-random patterns**…
- **Lenticular images and lenses** for 3d display or plenoptic acquisition…
- Use of the **chromatic aberration** to address depth ambiguity in defocus (works of Pauline Trouvé et al, 2013)
- **Translation of the focal plane** during acquisition to increase depth field (works of Hajime Nagahara, 2008)
- Applications of **event-based cameras**: works of ETH / Dynamic Vision Sensor or ISIR / Prophesee…
- **Coded aperture and vision of the octopus**: maximising the chromatic aberration for coloured vision (and more?) with monochromatic photoreceptors…