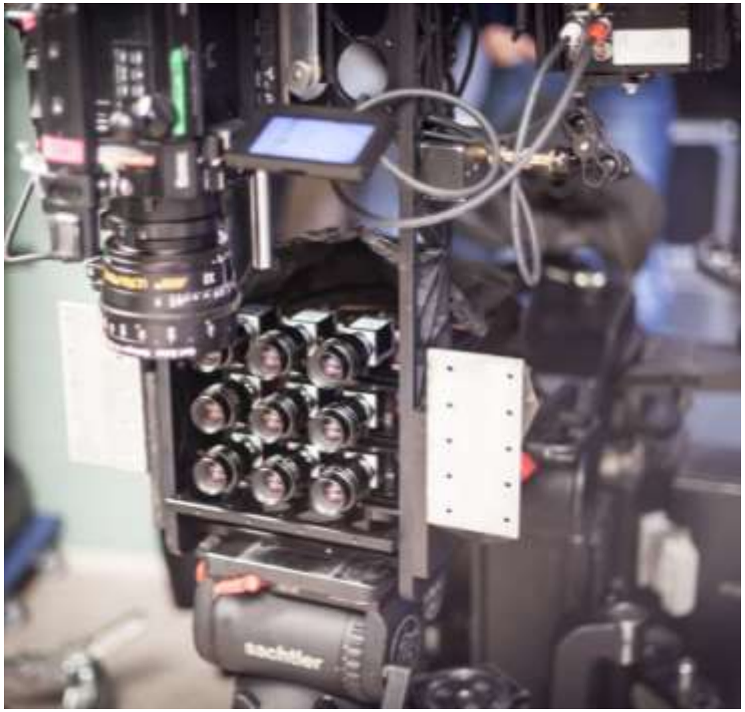
LIGHT FIELD CAPTURE FOR MEDIA PRODUCTION

Joachim Keinert, Fraunhofer IIS



Fraunhofer
IIS

TABLE OF CONTENT

TABLE OF CONTENT

Fraunhofer IIS

Coming Home – A test production

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot
Scene overview

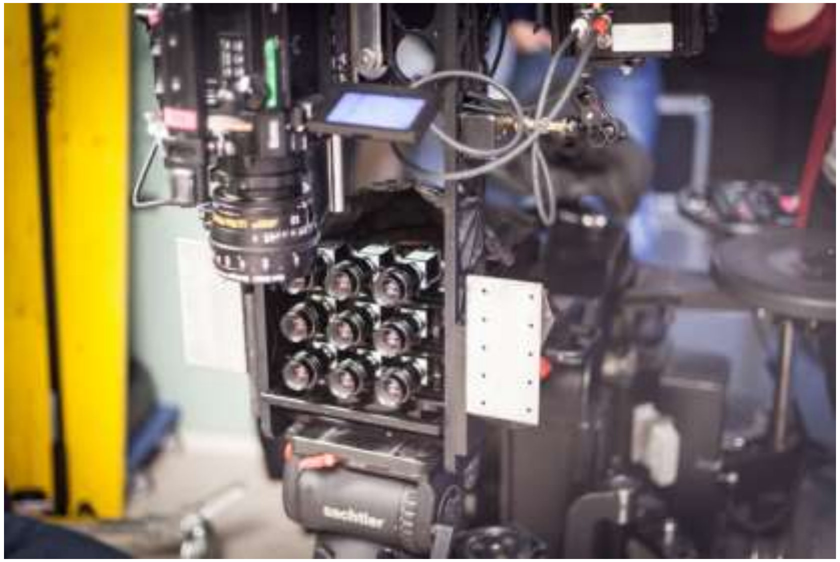
Coming Home – Some impressions from the shot
Many people involved

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot
Many people involved

Coming Home – Some impressions from the shot
Camera setup



HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot
Camera movement

Coming Home – Some impressions from the shot
Camera movement

Coming Home – Some impressions from the shot
Scene setup …

Coming Home – Some impressions from the shot … and focus pulling

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot … and focus pulling

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot …and focus pulling

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot
The challenge of lighting

HOCHSCHULE DER MEDIEN

Fraunhofer
IIS

Coming Home – Some impressions from the shot
Many tricks

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot
Many tricks

HOCHSCHULE DER MEDIEN

Fraunhofer IIS

Coming Home – Some impressions from the shot
Many tricks

Coming home – Making off

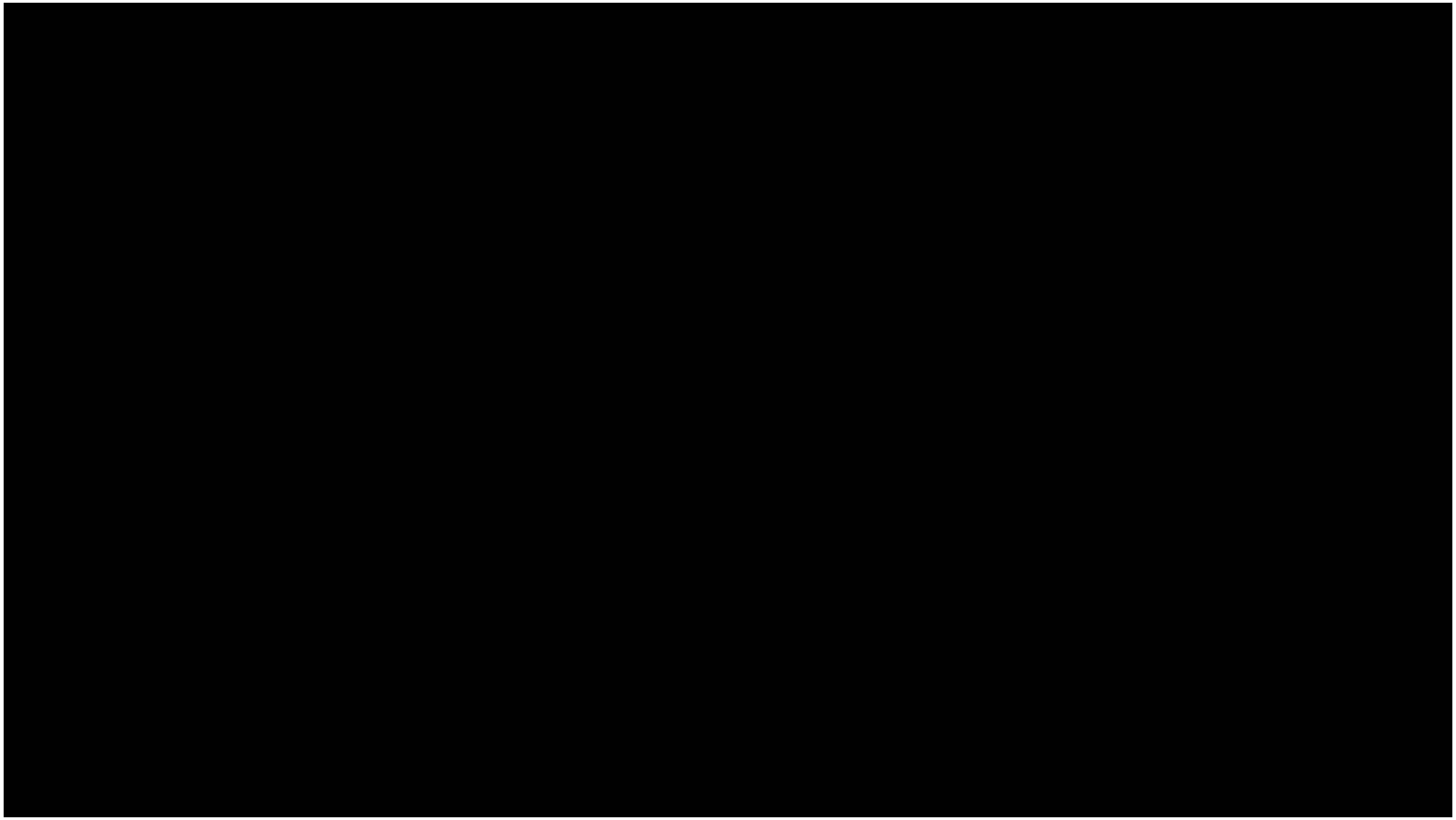
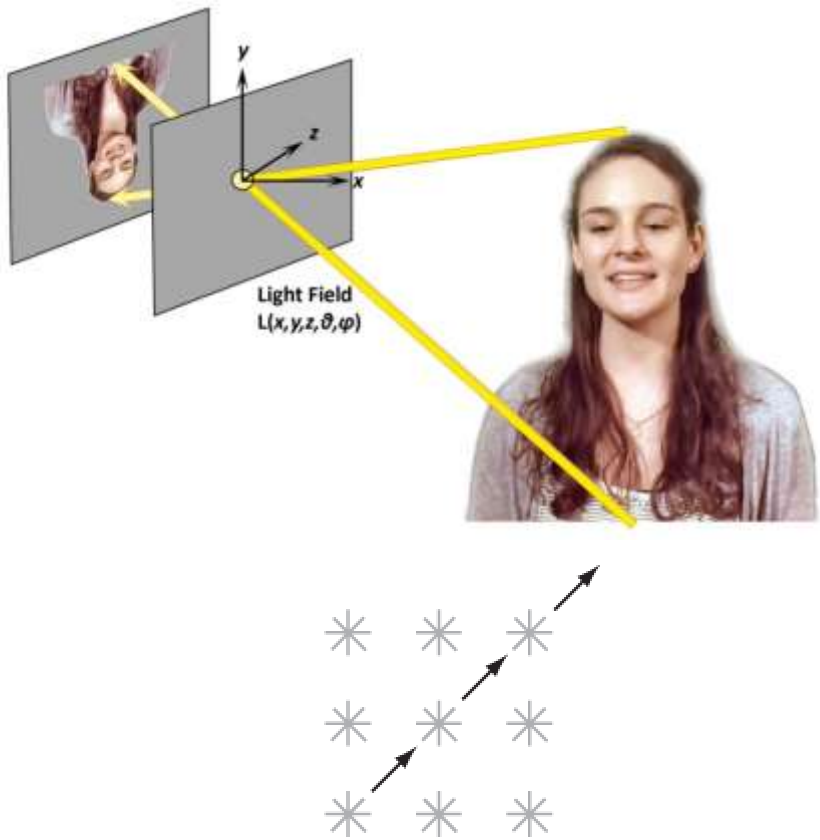


TABLE OF CONTENT

Fraunhofer IIS

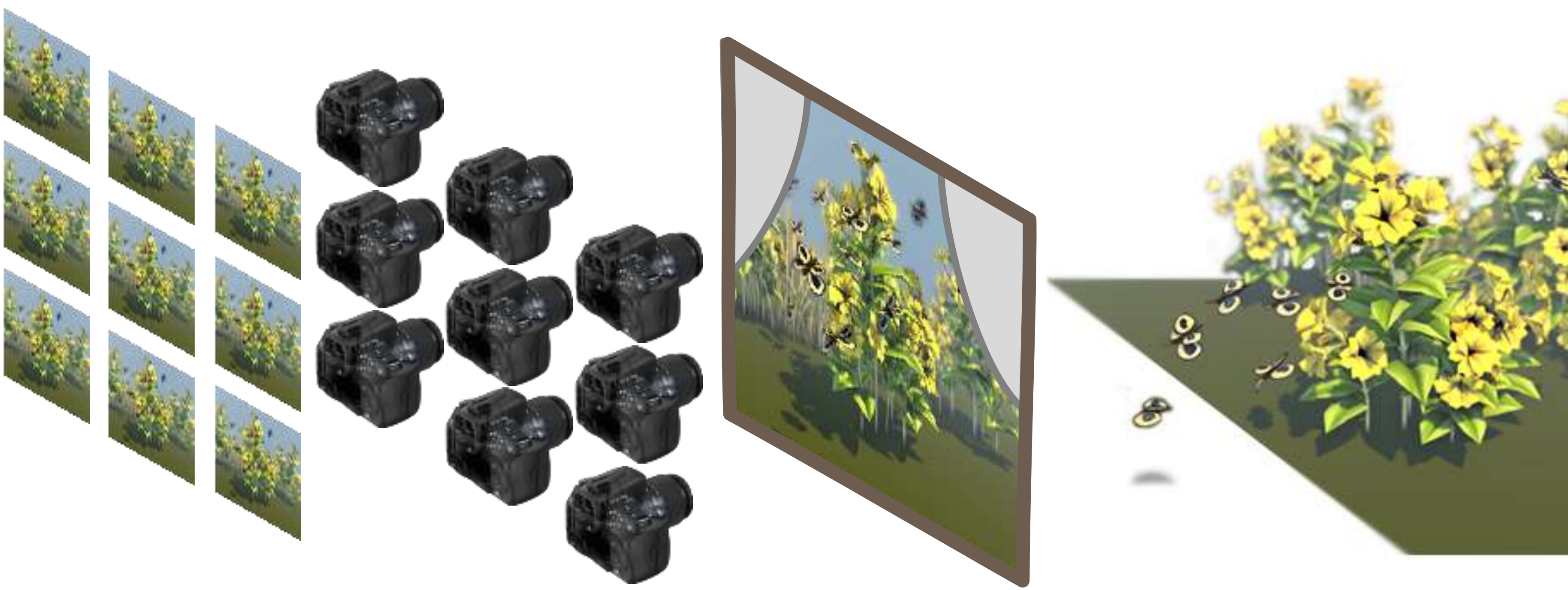
Definition of light field



- Intensity of all light rays in space
- 7D function $P_f(x, y, z, \theta, \phi, \lambda, t)$
  - spatial position (x, y, z)
  - viewing direction (θ, ϕ)
  - wavelength (λ)
  - time (t)
- Highly redundant when ignoring
  - Attenuation
  - Occulusions

Fraunhofer IIS

4D light field
Looking like through a window …



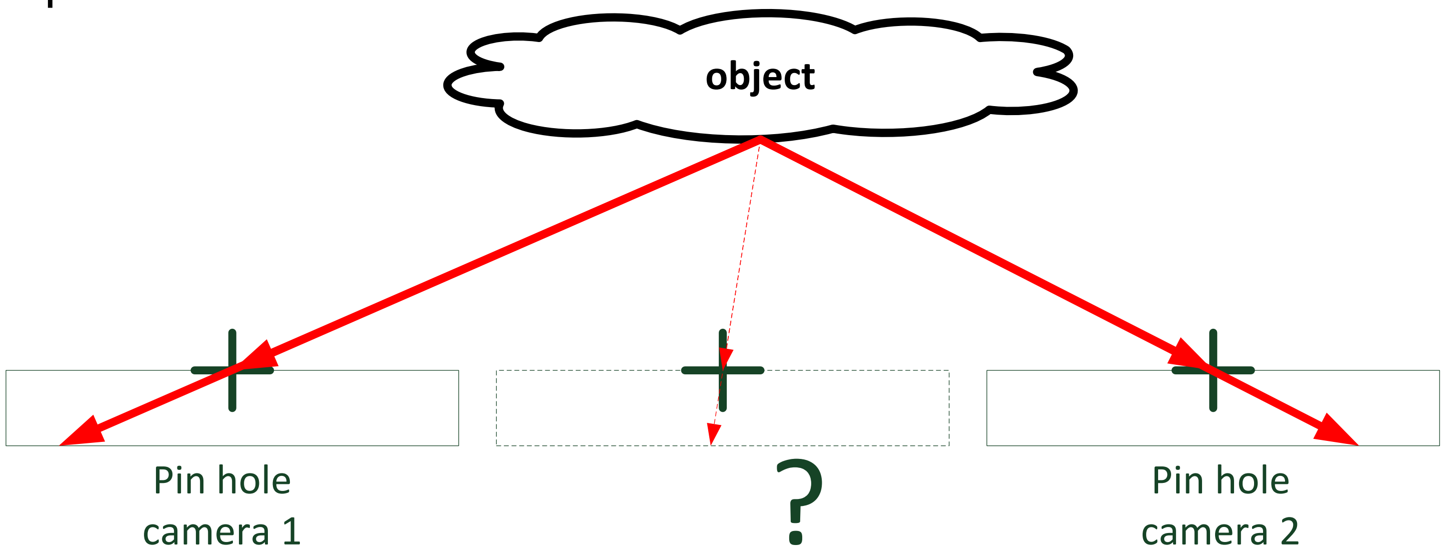
Sensor Projection Window World

Goal: Capture large scale parallax

Possible applications

Multi-camera capture
Problem formulation

- Sampling a 4D light field theoretically requires infinitely many cameras
- In practice, only a sparse sampling is possible
- Need to interpolate missing images
- Availability of depth permits determine origin of ray for correct interpolation



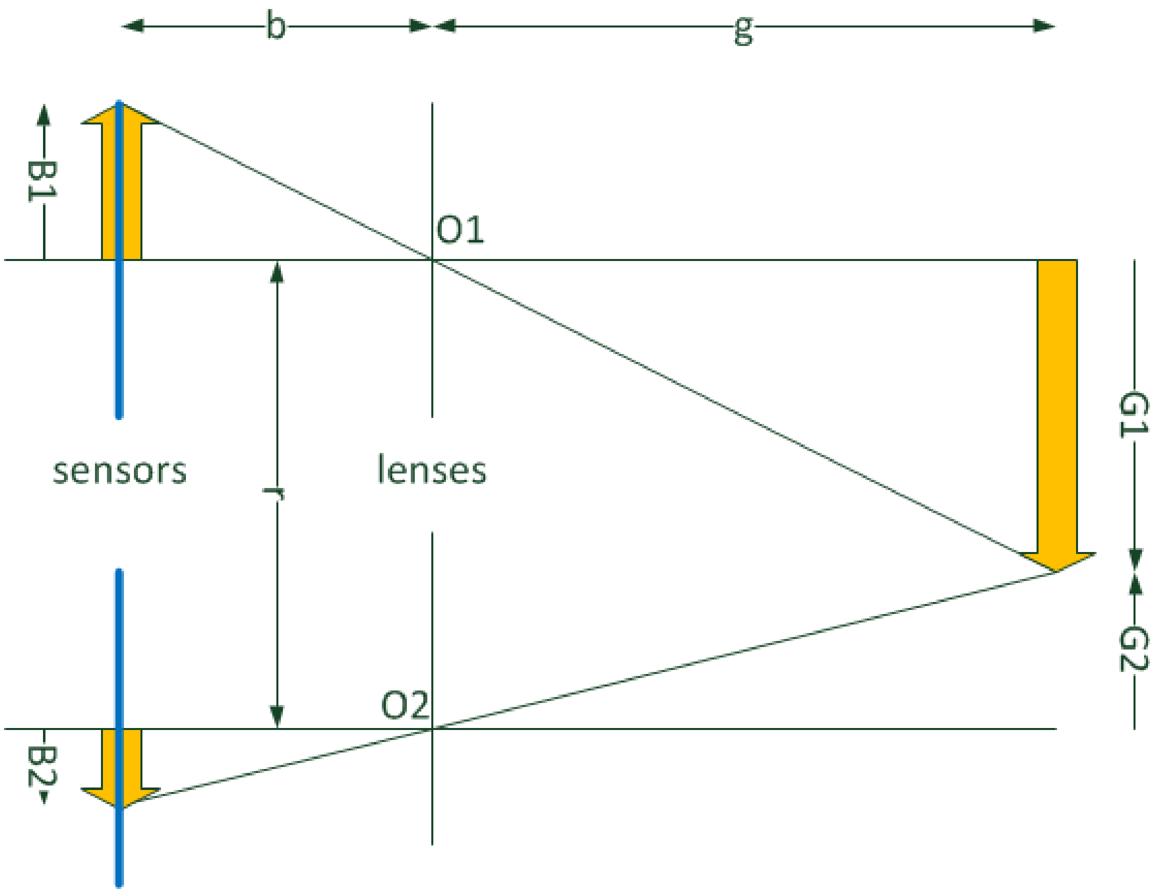
Fraunhofer IIS

TABLE OF CONTENT

Disparity and depth

- Sensors centred in optical axis
- $d^s = B1 - B2$
- $g \approx \frac{f}{d^s} \cdot r$
- $d^s \approx \frac{f}{g} \cdot r$
- In pixels

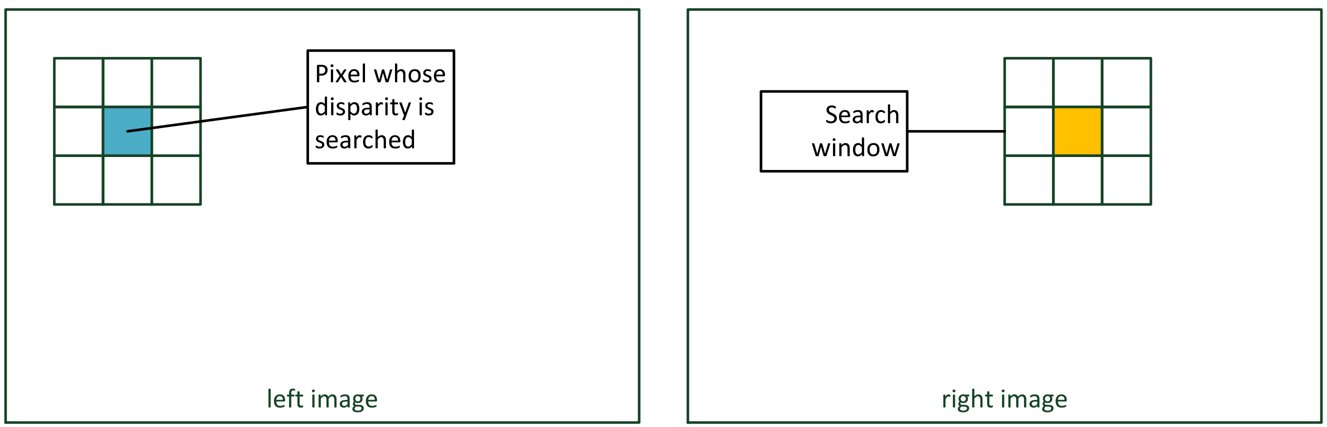
$$d^s_{pixel} = \frac{f}{g} \cdot \frac{r}{s}$$



Fraunhofer IIS

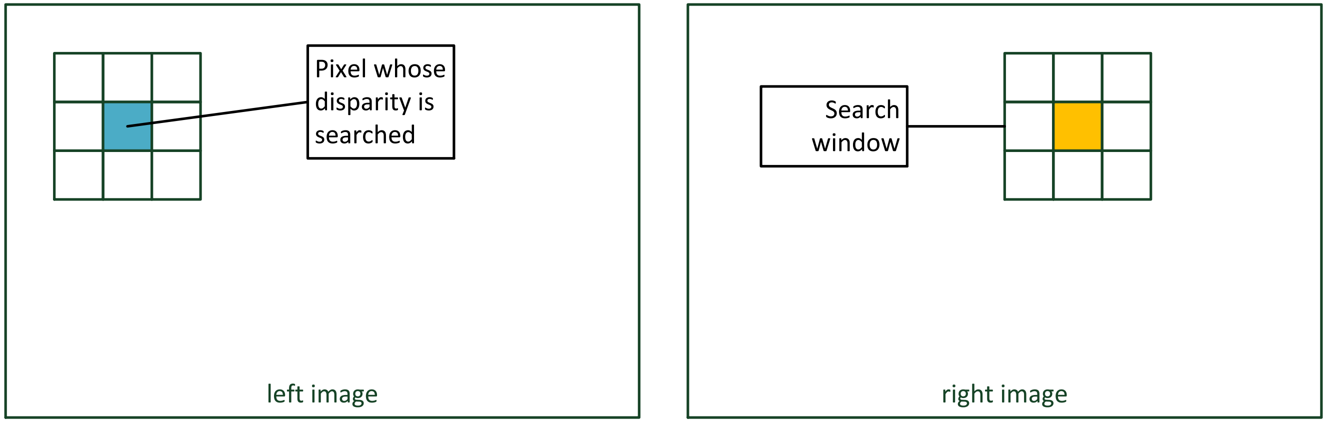
Dimensioning the array system

- Computation of disparity requires determination of correspondences
- Pixel based correspondences not possible
  - Ambiguous
  - Noise
- Correspondences are determined based on pixel regions



Dimensioning the array system

- Correspondences are determined based on pixel regions
- Disparity from the most similar search window
- Ideally, within the search window only limited depth variations
  - Otherwise disparity for pixels within window different
  - Windows are not similar anymore



Fraunhofer IIS

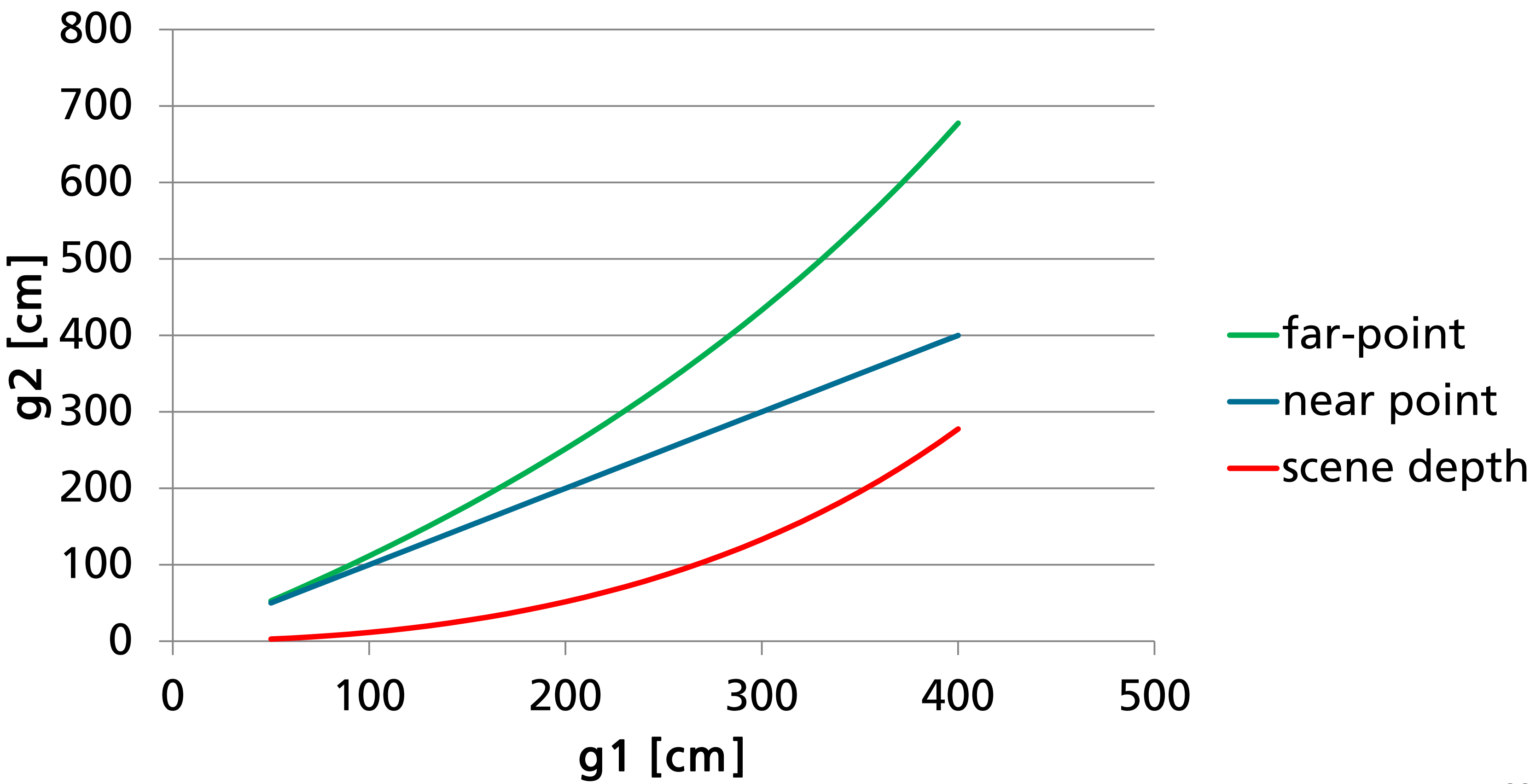
Dimensioning the array system

- Ideally, within the search window only limited depth variations
- Most critical is a window that contains both
  - Objects from the very front of the scene
  - Objects from the very back of the scene
- Disparity change (parallax):

$$\Delta \mathrm{d} = \frac{f \cdot r}{s} \cdot \left(\frac{1}{g_1} - \frac{1}{g_2} \right)$$

- Rule of thumb
  - $\Delta d = 0.03 \cdot w$
  - w: Number of horizontal pixels

Condition on far point for given near point
f=50mm, s=5µm, r=6cm, w=2048



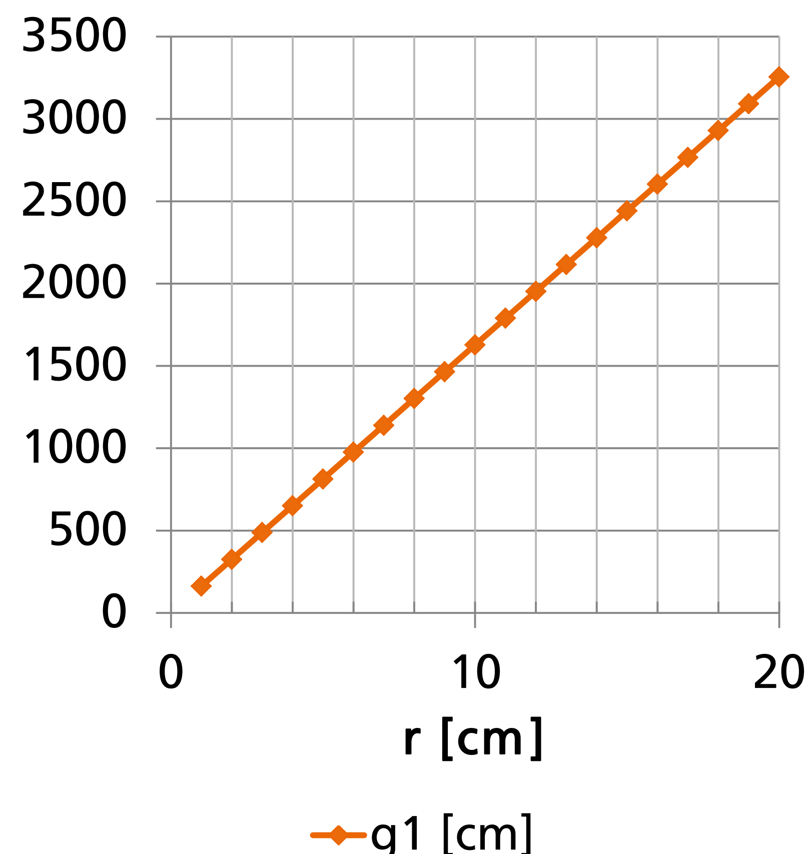
Fraunhofer IIS

Near point for far point at infinity f=50mm, s=5µm, w=2048

$$0.03 \cdot \mathrm{w} = \Delta \mathrm{d} = \frac{f \cdot r}{s} \cdot \left(\frac{1}{g_1} - \frac{1}{g_2}\right)$$

- $g_2 \to \infty$

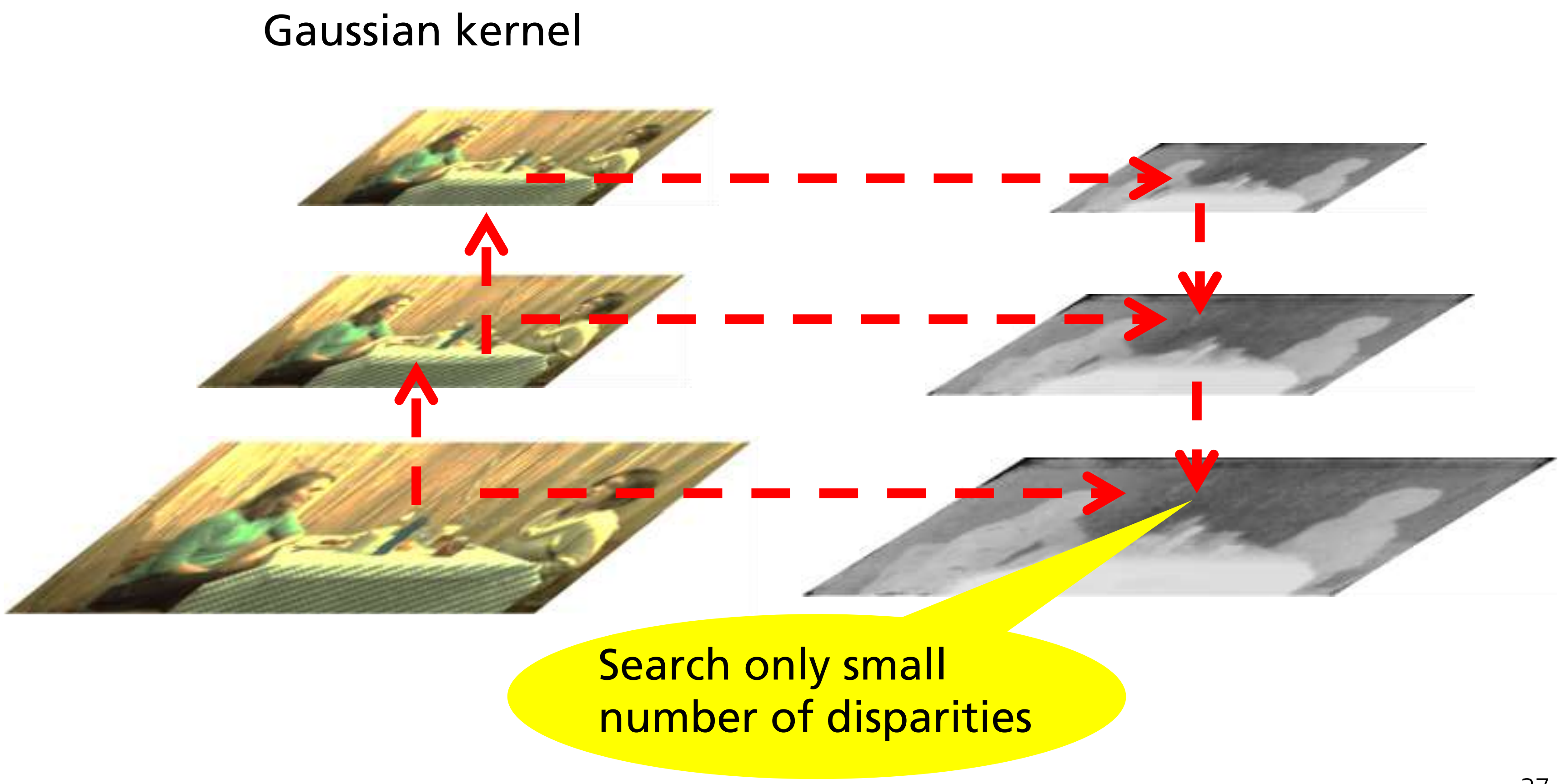
$$g_1 = \frac{f \cdot r}{s \cdot 0.03 \cdot w}$$



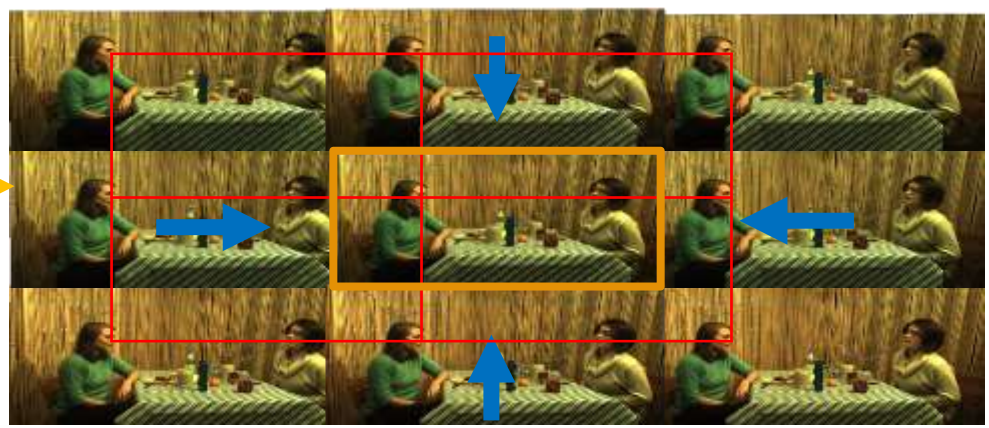
Fraunhofer IIS

TABLE OF CONTENT

Multi resolution approach



Pipeline Overview (1)

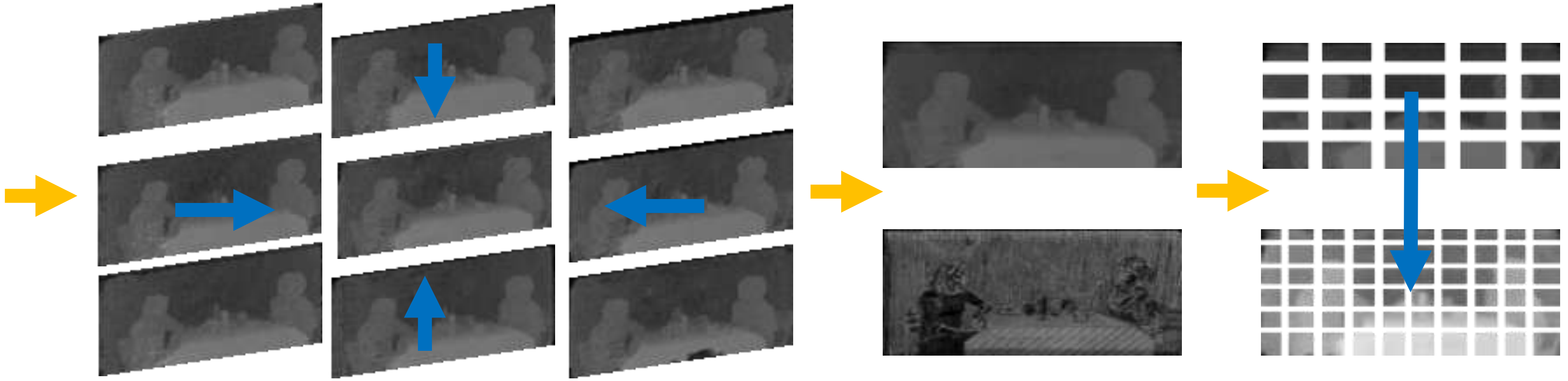


Disparity

Confidence

Capture Rectification Matching Per view result

Pipeline Overview (2)



Consolidation

Per view result

Upsampling

max planck institut informatik

Fraunhofer IIS

Census Transform

- Illumination invariant
- Encodes local brightness changes into a single integer variable
  - Every bit corresponds to a window position
- Applied census transform size is 5x5

Census transform

127	127	129
126	X	129
127	131	131

$X \geq Y$

X = 128

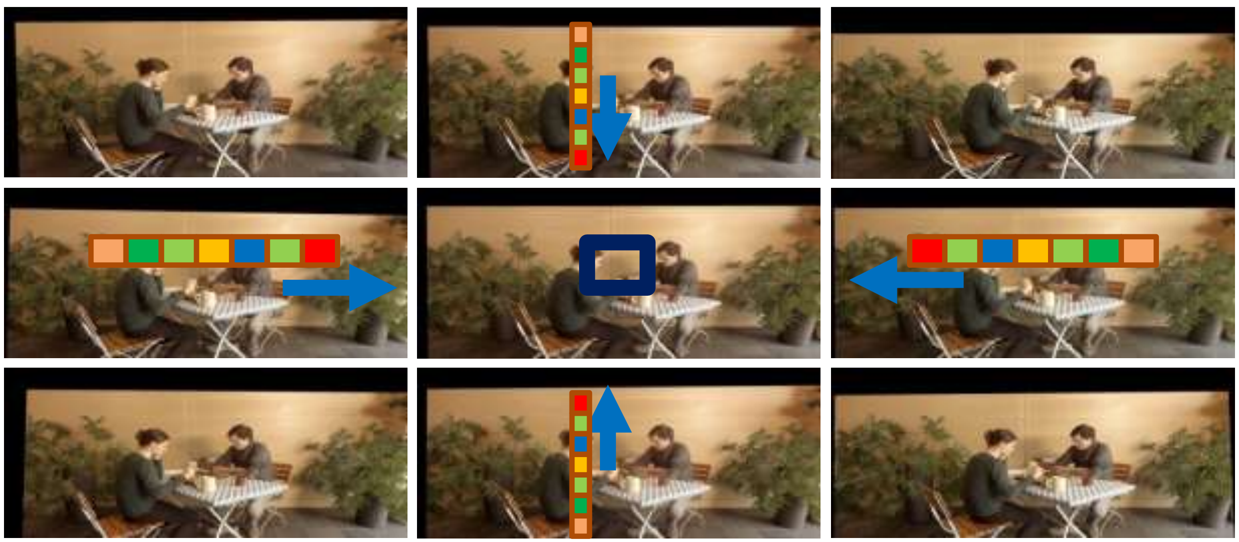
1	1	0	1	1	0	1	0	0

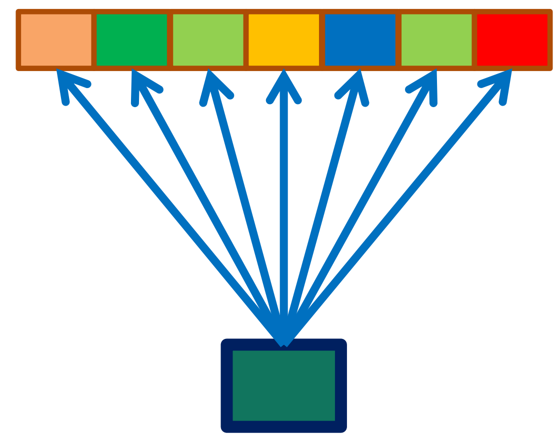
Multi-camera matching

- Relation between disparity d^s and depth g (when sensor centred in optical axis)

$$d^s \approx \frac{1}{g} \cdot f \cdot r$$

- Disparity between different camera pairs proportional to base line





Multi-camera cost computation

- Costs for a single camera pair
  - Number of different bits in the census transform output
- Costs for multiple camera pairs
  - Mean of all single camera pair costs
- Computation of disparity
  - Those disparity that have minimum costs

Census transform

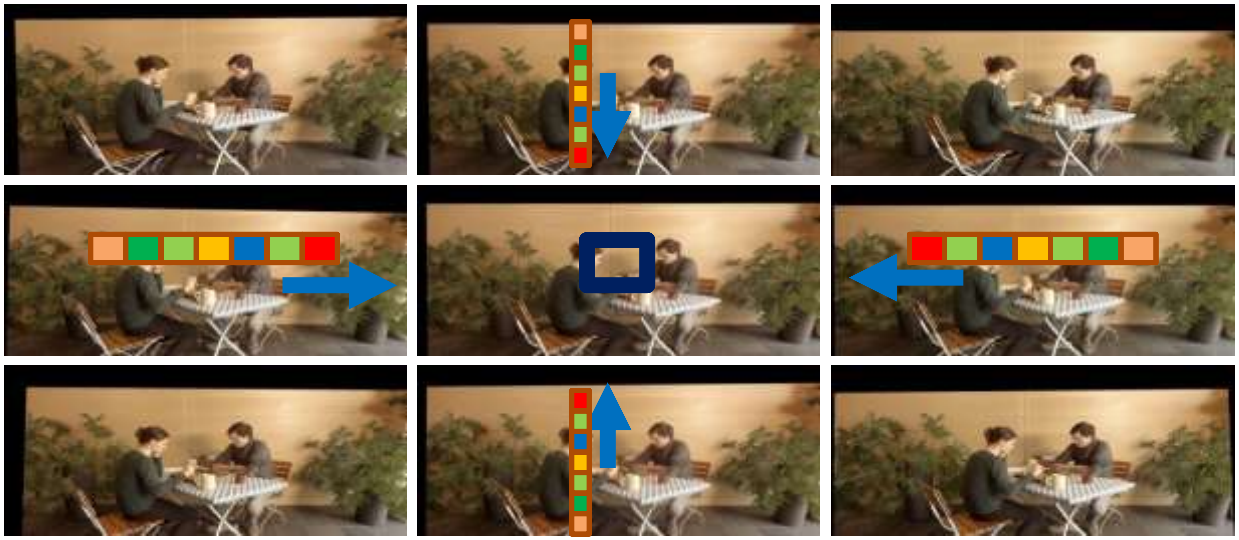
127	**127**	**129**
126	X	129
127	131	131

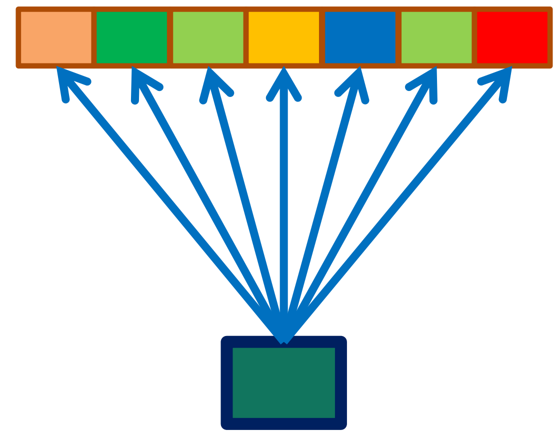
$X \geq Y$

X = 128

1	1	0	1	1	0	1	0	0

Matching confidence



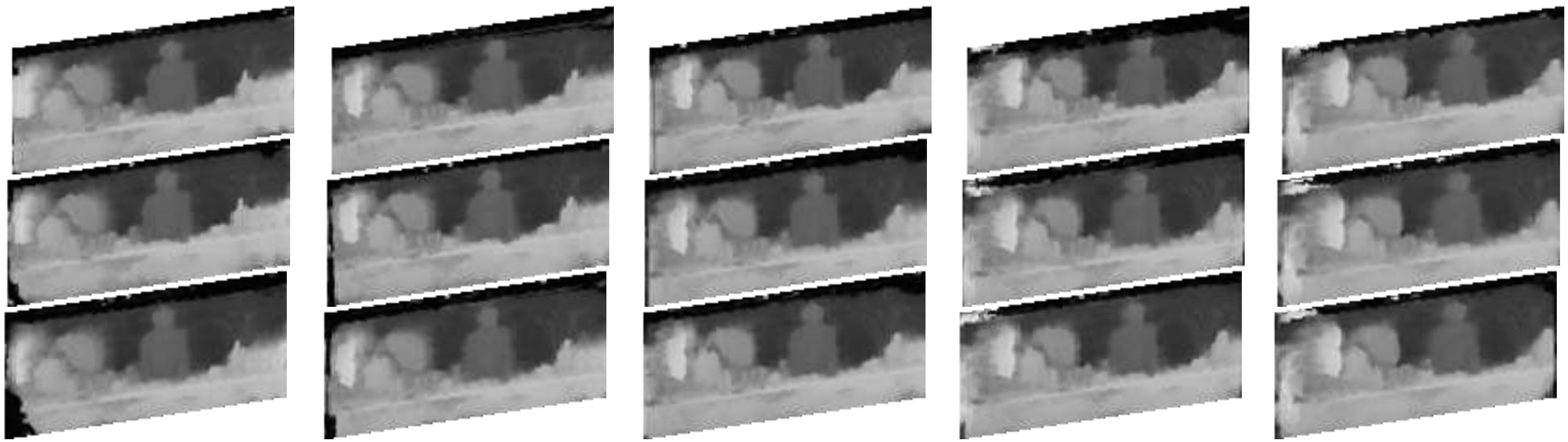


- How sure is decision for view i ?
- $\overline{C}_i(\vec{x}) = 1 - \frac{1}{1+\sigma_m \cdot h_i(\vec{x})}$
- $h_i = |average\ cost(tested\ disparities) - best\ cost|^{1/2}$

max planck institut informatik

Fraunhofer IIS

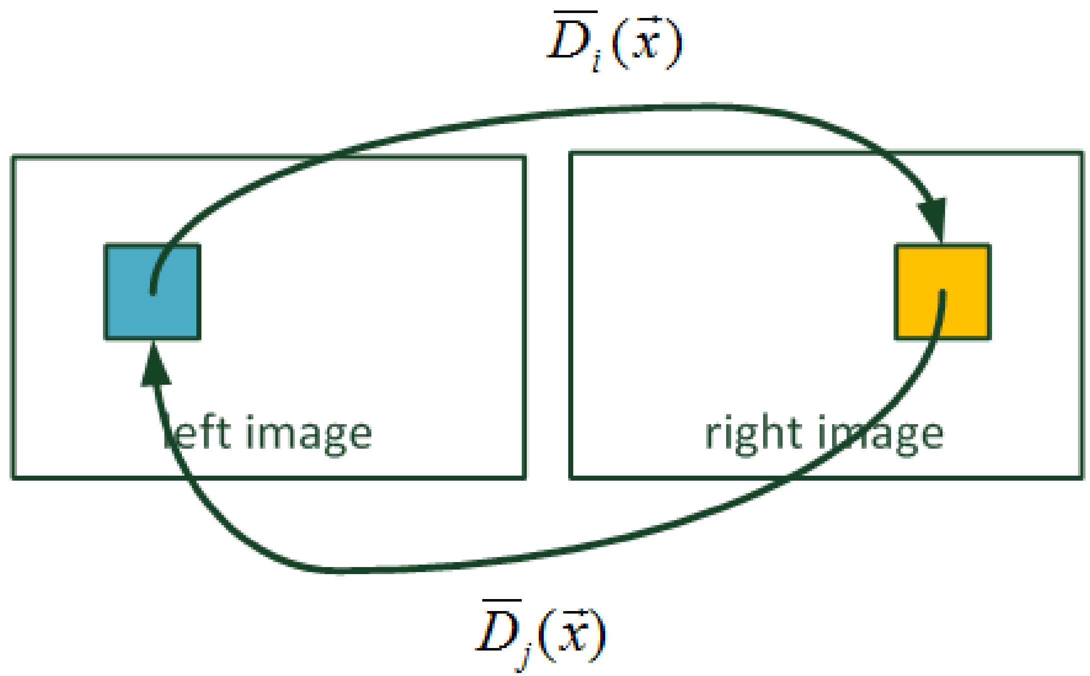
Consolidation of different views



- Disparity maps of different views might contradict
  - noise
  - occlusions
  - Specularities
- Contradicting disparities should have low confidence

Consolidation of different views
Principle

- Disparity $\overline{D_i}(\vec{x})$ obtained for view i
- Disparity $\overline{D_j}(\vec{x})$ obtained for view j



Consolidation of different views
Computation of overall confidence

- Input
  - Matching confidence per view i: $\overline{C_i}(\vec{x}) = 1 - \frac{1}{1+\sigma_m \cdot h_i(\vec{x})}$
  - Disparity $\overline{D_i}(\vec{x})$ obtained for view i
- Consolidated confidence for view i at pixel $\vec{x}$

$$C_i(\vec{x}) = \sum_{i_2} \hat{C}_{i,i_2}(\vec{x})$$

- Confidence of view i relative to view i_2 (for equidistant cameras)

$$\hat{C}_{i,i_2}(\vec{x}) = \frac{\overline{C_i}\left(\overrightarrow{x'}\right)}{1 + \sigma_r \cdot \underbrace{\left|\overline{D_i}(\vec{x}) - \overline{D_{i_2}}\left(\overrightarrow{x'}\right)\right|}_{reprojection\ error}}, \overrightarrow{x'} = \vec{x} + \overline{D_i}(\vec{x})$$

Consolidation of different views
Computation of overall confidence

- Confidence of view i relative to view i_2

$$\hat{C}_{i,i_2}(\vec{x}) = \frac{\overline{C_i}\left(\overrightarrow{x'}\right)}{1 + \sigma_r \cdot \underbrace{\left|\overline{D_i}(\vec{x}) - \overline{D_{i_2}}\left(\overrightarrow{x'}\right)\right|}_{reprojection\ error}}$$

- No reprojection error
  - Neighbour view i_2 confirms “hypothesis” of view i
  - Increase confidence term
- Large reprojection error
  - Neighbour view i_2 does not confirm “hypothesis” of view i
  - Do not increase confidence term
- By these means wrong disparity values will obtain low confidence values

Consolidation of different views
Disparity combination from different views

- Consolidated disparity for view i at pixel $\vec{x}$

$$D_i(\vec{x}) = \sum_{i_2} \left(\overline{D_{i_2}}\left(\overrightarrow{x'}\right) \cdot \frac{\hat{C}_{i,i_2}(\vec{x})}{C_i(\vec{x})} \right), \overrightarrow{x'} = \vec{x} + \overline{D_i}(\vec{x})$$

- Improve disparities that are slightly off due to noise (low confidence!)
- Will not help for disparities that are wrong anyhow

Search space reduction by multi-resolution approach

- Upscaling needs to be border sensitive!
- Also ignore disparities with low confidence

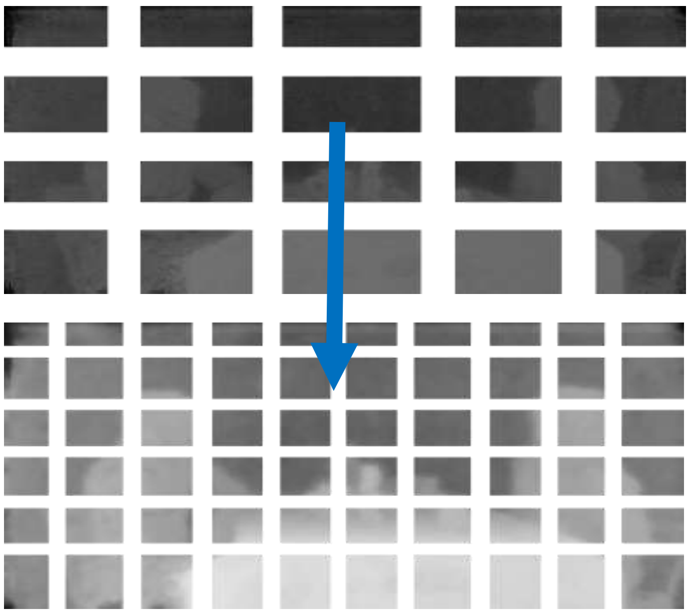
create approximation by upscaling

Only search in a small interval centred in the approximation value

Upsampling

$D_i^{(j)}$

$C_i^{(j)}$



$D_i^{(j)}$

$\uparrow D_i^{(j)}$

Edge aware upsampling

- What disparities should be taken to the next level?
  - High resolution target pixel x
  - Low resolution source pixel y

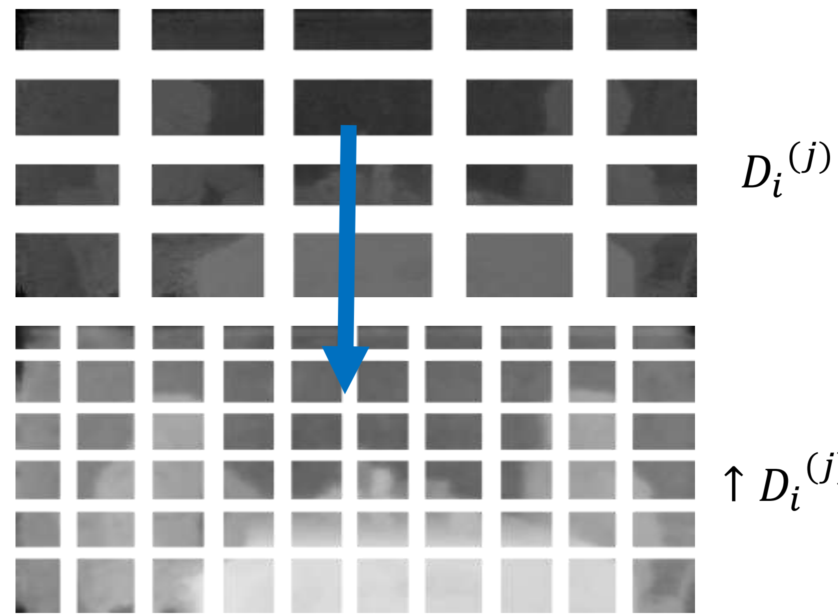
$$D_i^{(j)}(x) = \frac{1}{Z} \cdot \sum_{y \in N_i(x)} w(x,y) \cdot D_i^{(j)}(y)$$

$$w(x,y) = C_i^{(j)}(y)\, w_s(x,y)\, w_a(x,y)$$

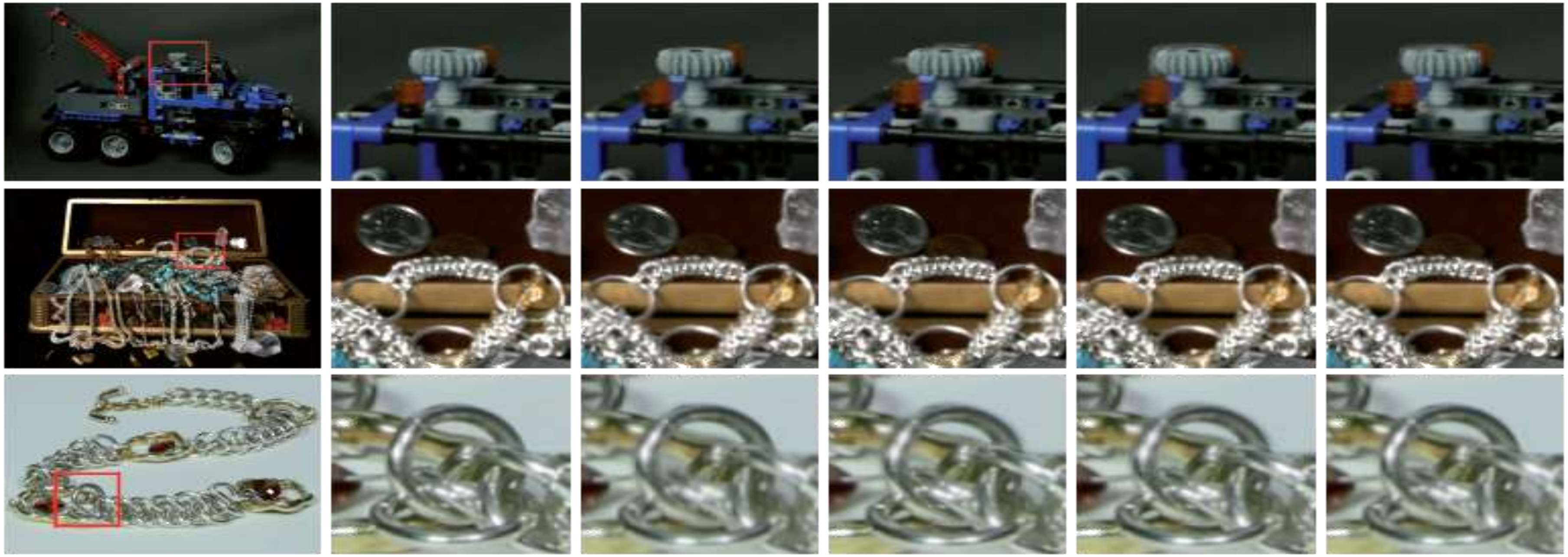
Confidence

Proximity

Appearance

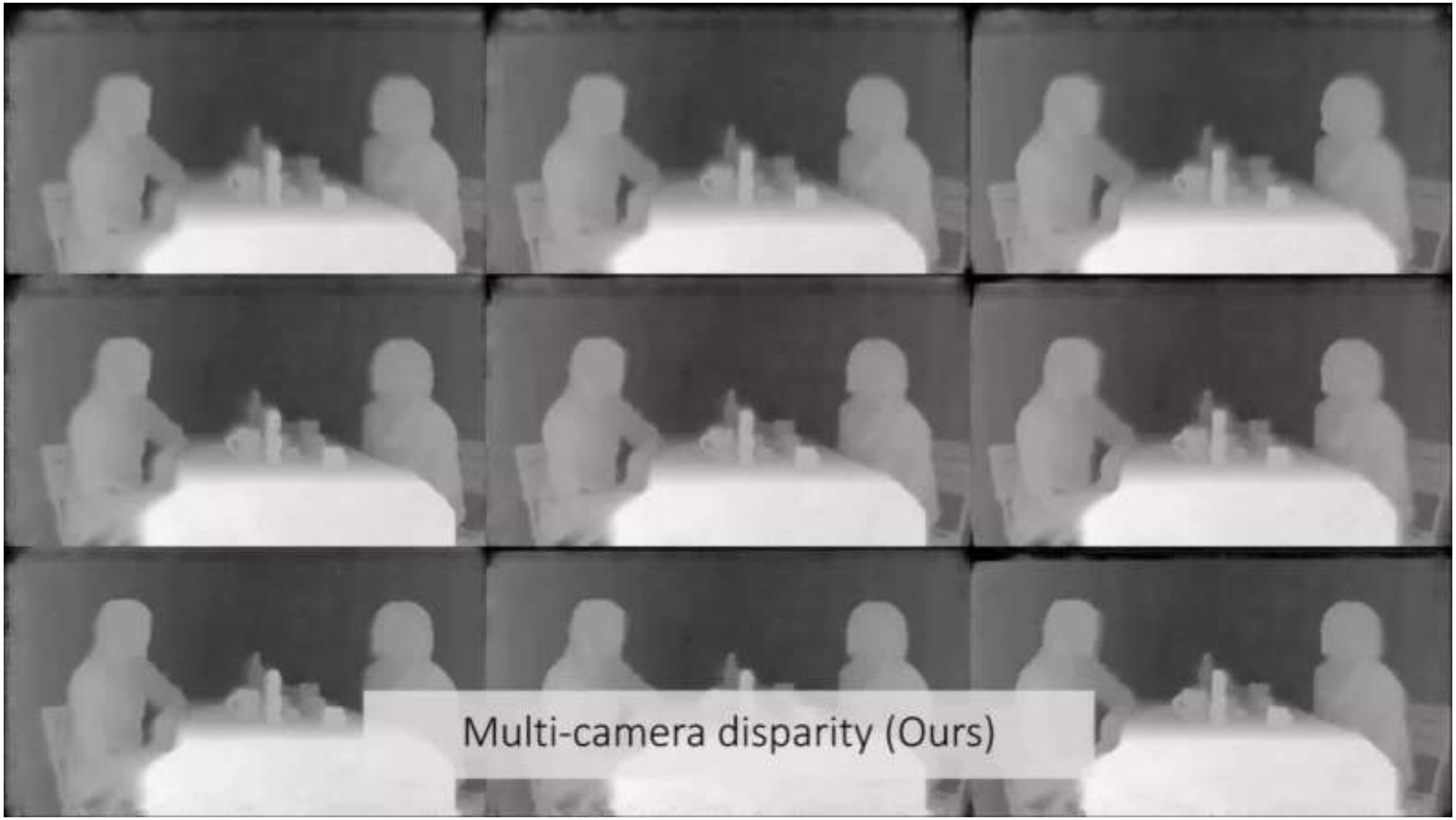


Dense light field results



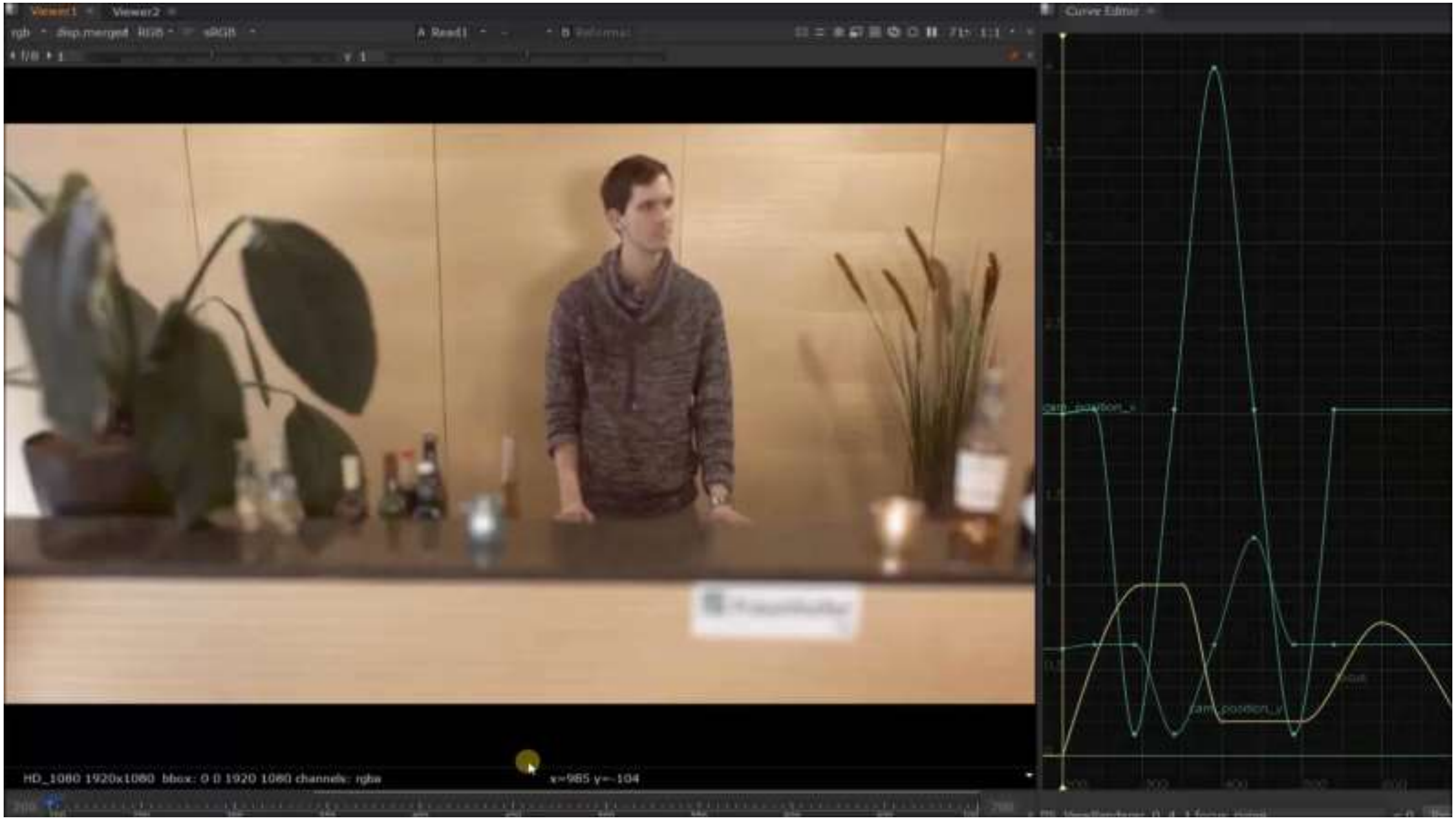
SGM | CVF | Wang | Our method

Sparse light field results



Depth based view synthesis

Light Field Media Production



max planck institut informatik

Fraunhofer IIS

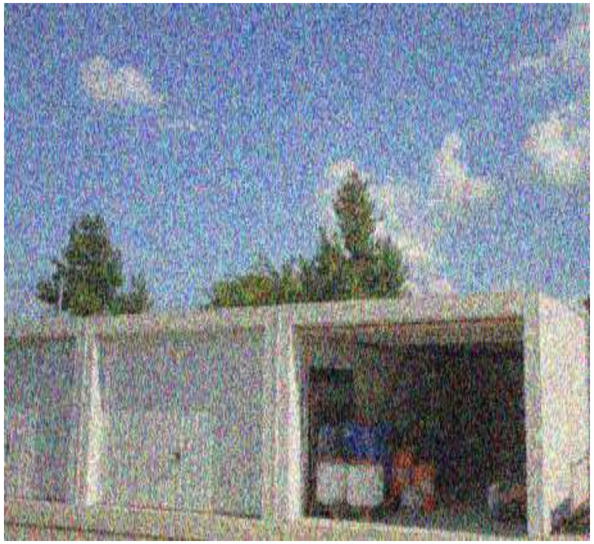
TABLE OF CONTENT

What are the properties of a high quality camera?

- **Low noise**
- **High dynamic range**
- High colour fidelity
- No geometric distortions
- No artefacts
  - Rolling shutter artefacts
  - Defect pixels
  - Blooming
  - …

Fraunhofer IIS

What are the properties of a high quality camera? Low noise



What are the properties of a high quality camera? Dynamic range

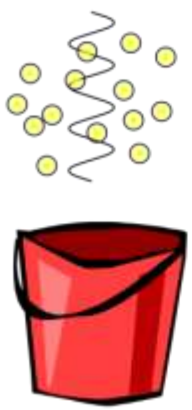
- Low dynamic range capture

- High dynamic range capture

What are the properties of a high quality camera? Dynamic range

- Bright objects

many photons



large bucket

- Dark objects

few photons

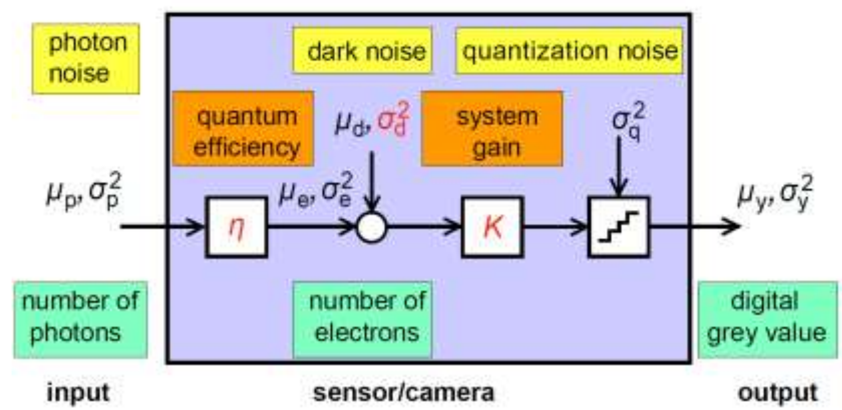
difficult to count

large bucket

Fraunhofer IIS

Model of a digital camera for systematic analysis

- μ_p: Mean number of photons hitting area A of single pixel during exposure time
- μ_e: Mean number of generated electric charges
- $\mu_e = \eta(\lambda) \cdot \mu_p$
  - Quantum efficiency



Fraunhofer IIS

Model of a digital camera for systematic analysis

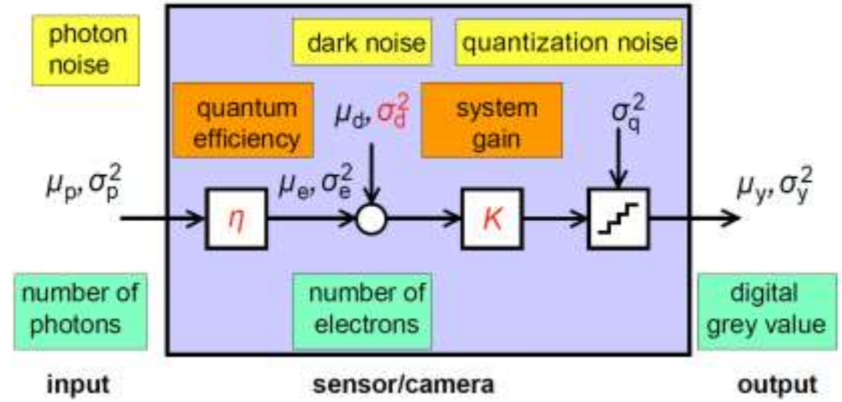
- Signal to noise ratio

$$SNR = \frac{\mu_y - \mu_{y,dark}}{\sigma_y}$$

- Dark current

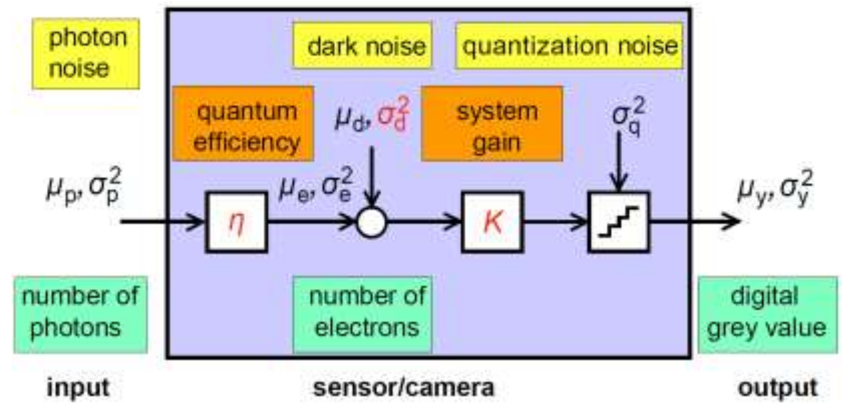
$$\mu_{y,dark} = \mu_d \cdot K$$

- In the following: $\mu_{y,dark} \approx 0$



Model of a digital camera
Noise sources

- Electronic noise σ_d^2
  - Transistors
  - Thermal noise
- Quantization noise σ_q^2
  - Signal dependent!
- Photon noise (shot noise) σ_e^2



Statistically independent noise sources

$$\sigma_y = K^2 \cdot \left(\sigma_d^2 + \sigma_e^2\right) + \sigma_q^2$$

Fraunhofer IIS

Model of a digital camera
Quantization noise

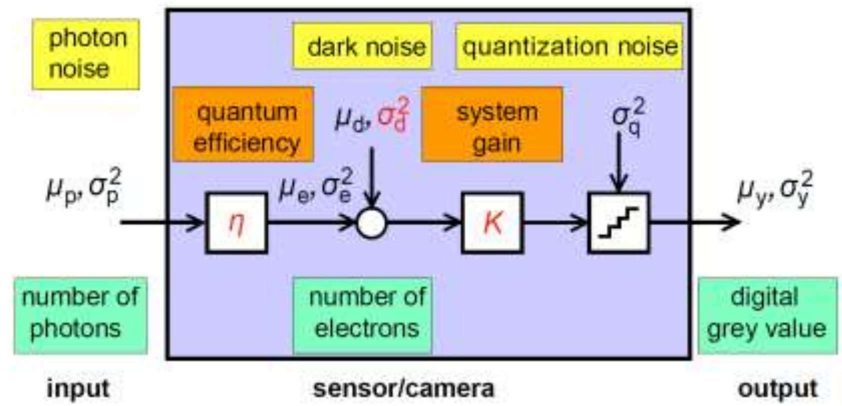
- $\sigma_q^2 = E\left[\left(X - E(X)\right)^2\right]$
- Consider first quantization bin
  - $E(X) \approx \frac{1}{2} DN$
- Quantization noise

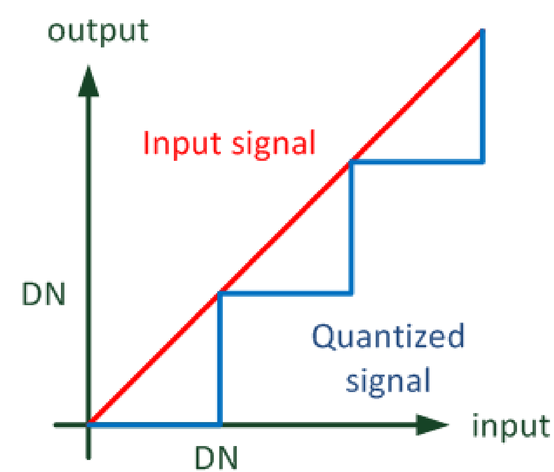
$$\sigma_q^2 \approx \frac{1}{DN} \cdot \int_0^{DN} \left(x - \frac{1}{2}DN\right)^2 dx$$

$$= \frac{1}{DN} \cdot \frac{1}{3} \cdot \left[\left(x - \frac{1}{2}DN\right)^3\right]_0^{DN}$$

$$= \frac{1}{12} DN^2$$

- Same for other quantization bins



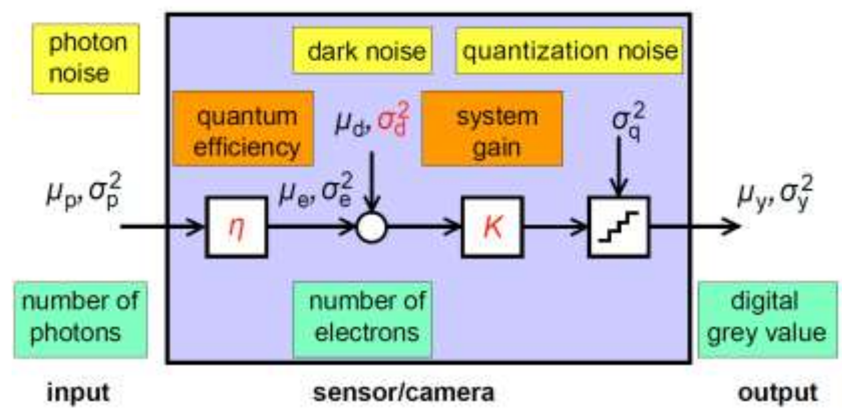


Fraunhofer IIS

Model of a digital camera
Photon shot noise

- Shot noise occurs in photon counting in optical devices
- Poisson process

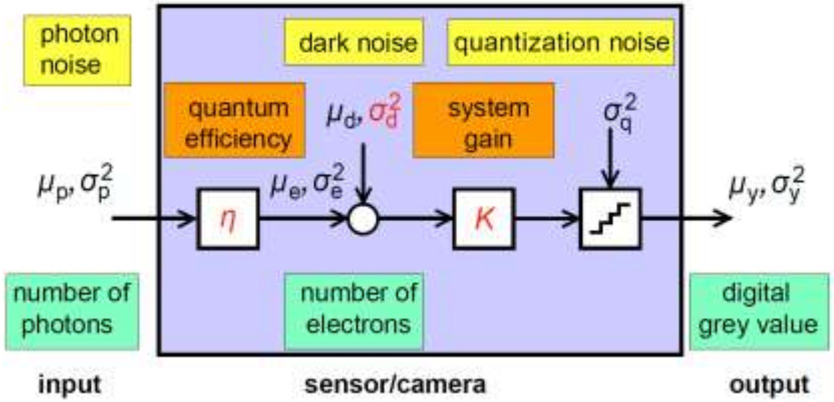
$$\sigma_e^2 = \mu_e$$



Fraunhofer IIS

Signal to noise ratio

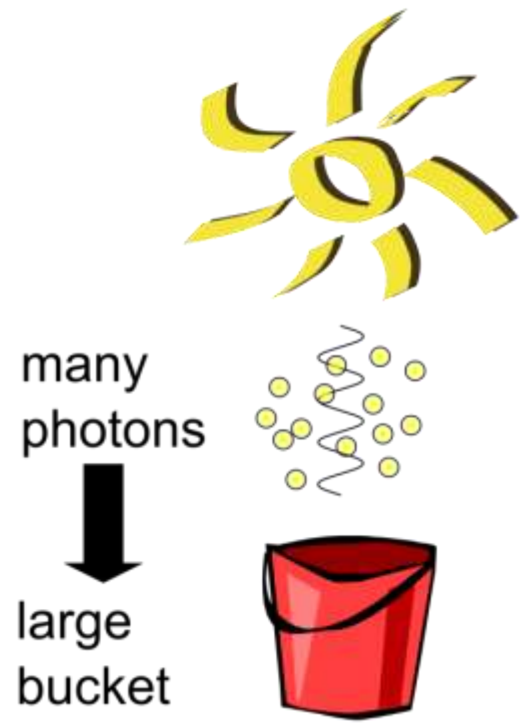
- $$SNR = \frac{\mu_y - \mu_{y,dark}}{\sigma_y} = \frac{\mu_e - \mu_d}{\sqrt{(\sigma_d^2 + \sigma_e^2) + \frac{\sigma_q^2}{K^2}}}$$
- $$SNR \approx \frac{\mu_y}{\sigma_y} = \frac{\mu_e}{\sqrt{(\sigma_d^2 + \mu_e) + \frac{\sigma_q^2}{K^2}}}$$
- $$SNR \approx \frac{\mu_y}{\sigma_y} = \frac{\eta \cdot \mu_p}{\sqrt{(\sigma_d^2 + \eta \cdot \mu_p) + \frac{\sigma_q^2}{K^2}}}$$
- $$SNR \approx \begin{cases} \sqrt{\eta \cdot \mu_p} & \eta \cdot \mu_p \gg \sigma_d^2 + \frac{\sigma_q^2}{K^2} \\ \frac{\eta \cdot \mu_p}{\sqrt{\sigma_d^2 + \frac{\sigma_q^2}{K^2}}} & \eta \cdot \mu_p \ll \sigma_d^2 + \frac{\sigma_q^2}{K^2} \end{cases}$$



Fraunhofer IIS

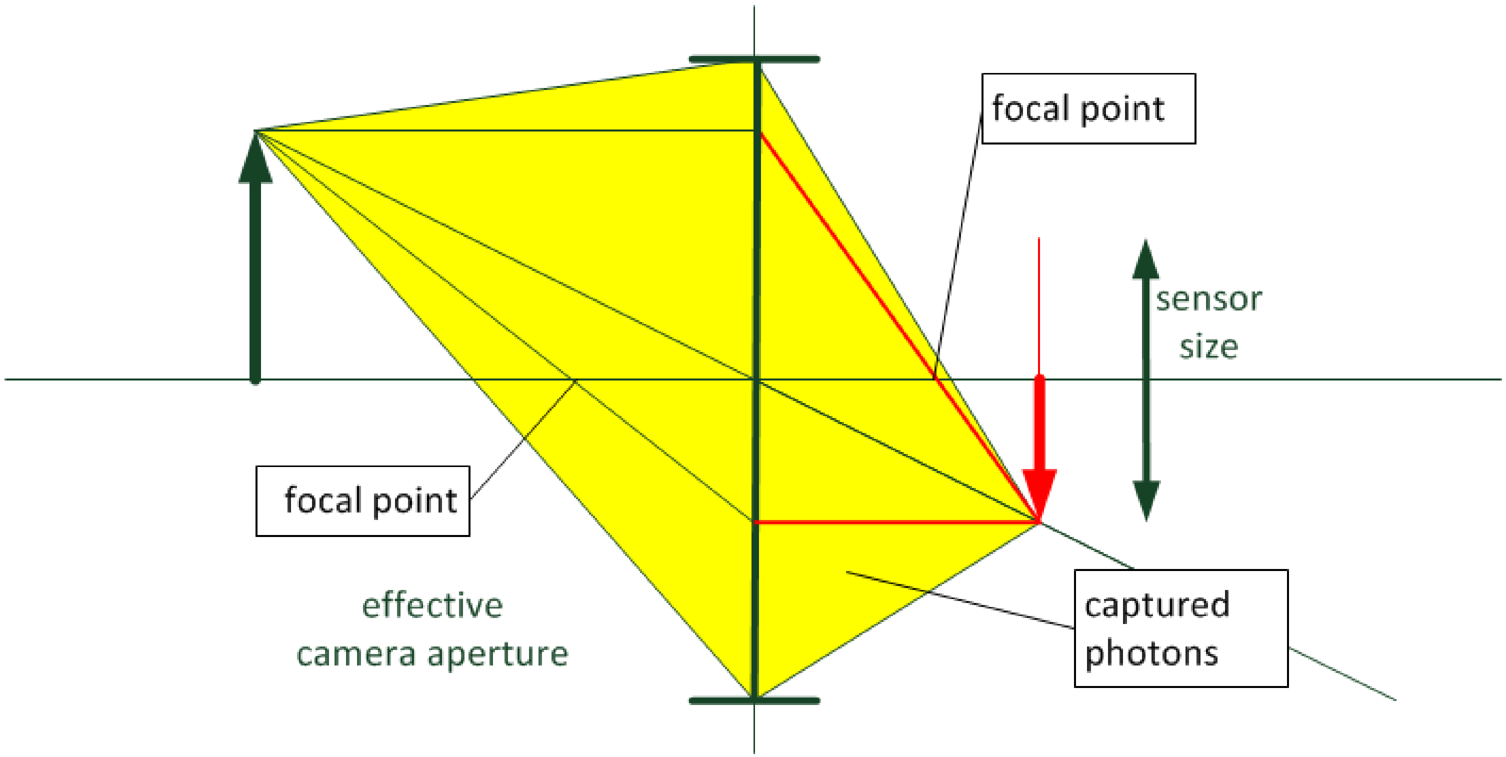
Quality considerations

- In good modern digital cameras, noise is dominated by photon shot noise
  - Only at the very lowest signal levels, electronic noise becomes a factor
- $SNR \approx \begin{cases} \sqrt{\eta \cdot \mu_p} & \eta \cdot \mu_p \gg \sigma_d^2 + \frac{\sigma_q^2}{K^2} \\ \frac{\eta \cdot \mu_p}{\sqrt{\sigma_d^2 + \frac{\sigma_q^2}{K^2}}} & \eta \cdot \mu_p \ll \sigma_d^2 + \frac{\sigma_q^2}{K^2} \end{cases}$
- $SNR \approx \sqrt{\eta \cdot \mu_p}$
- **Good image quality (low SNR) requires a lot of photons!**
- Solution: larger sensor?



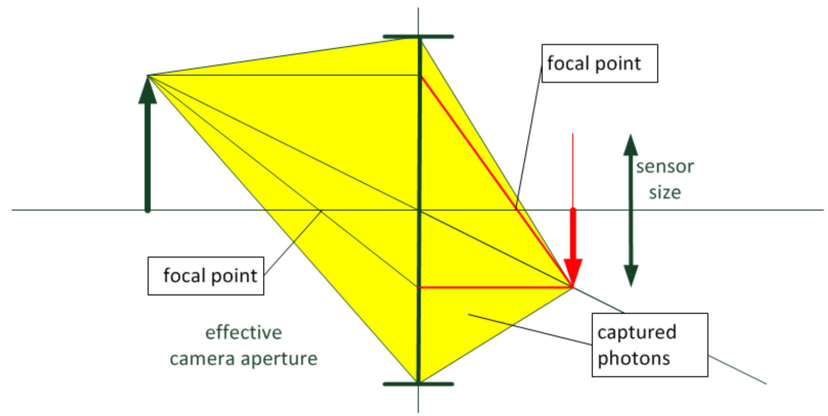
Fraunhofer IIS

Relation with optics
Small sensor

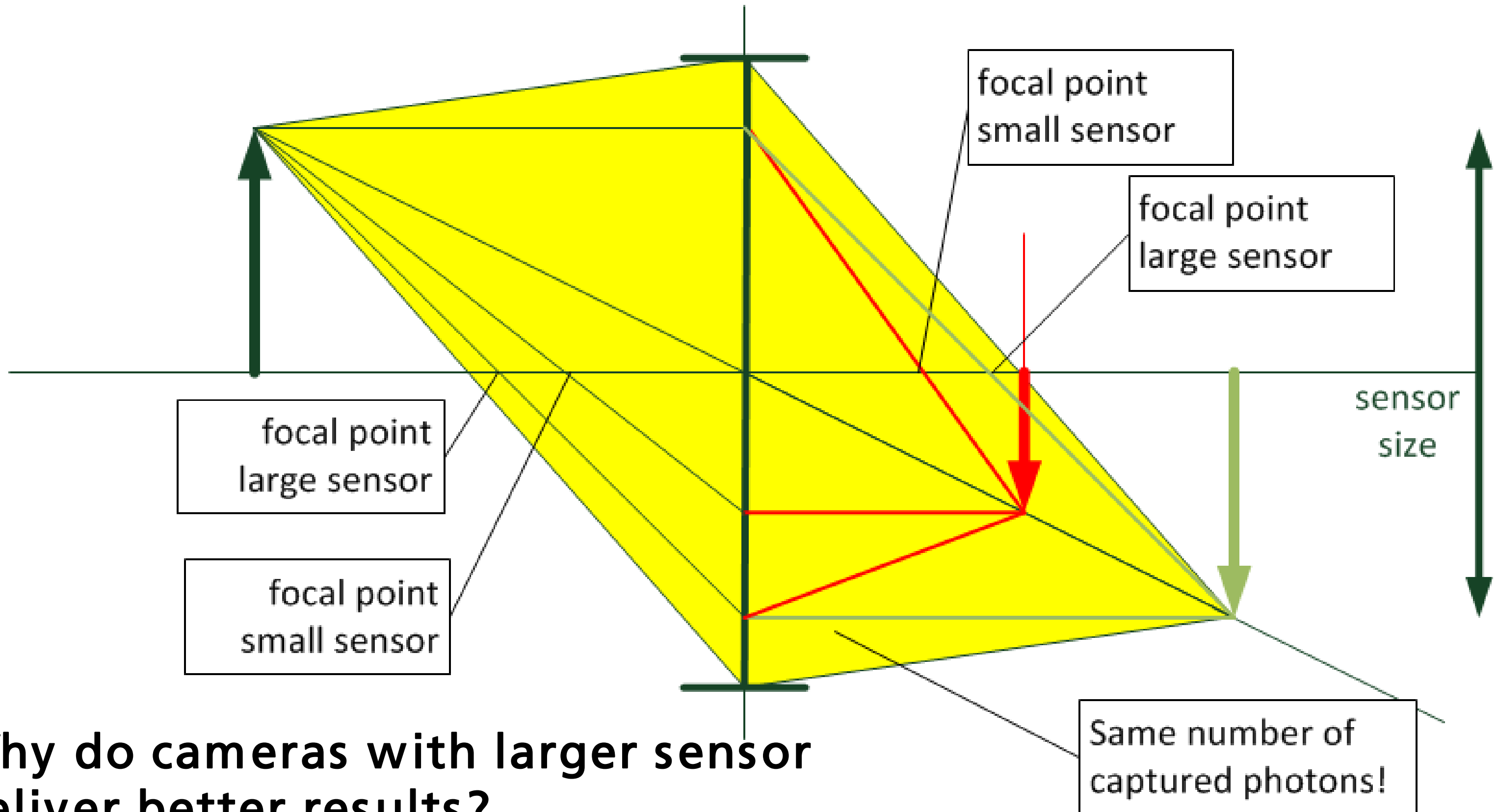


Relation with optics
Increasing sensor size

- Larger pixels would capture more photons
- Consequences
  - Resolution gets smaller
  - Field of view changes



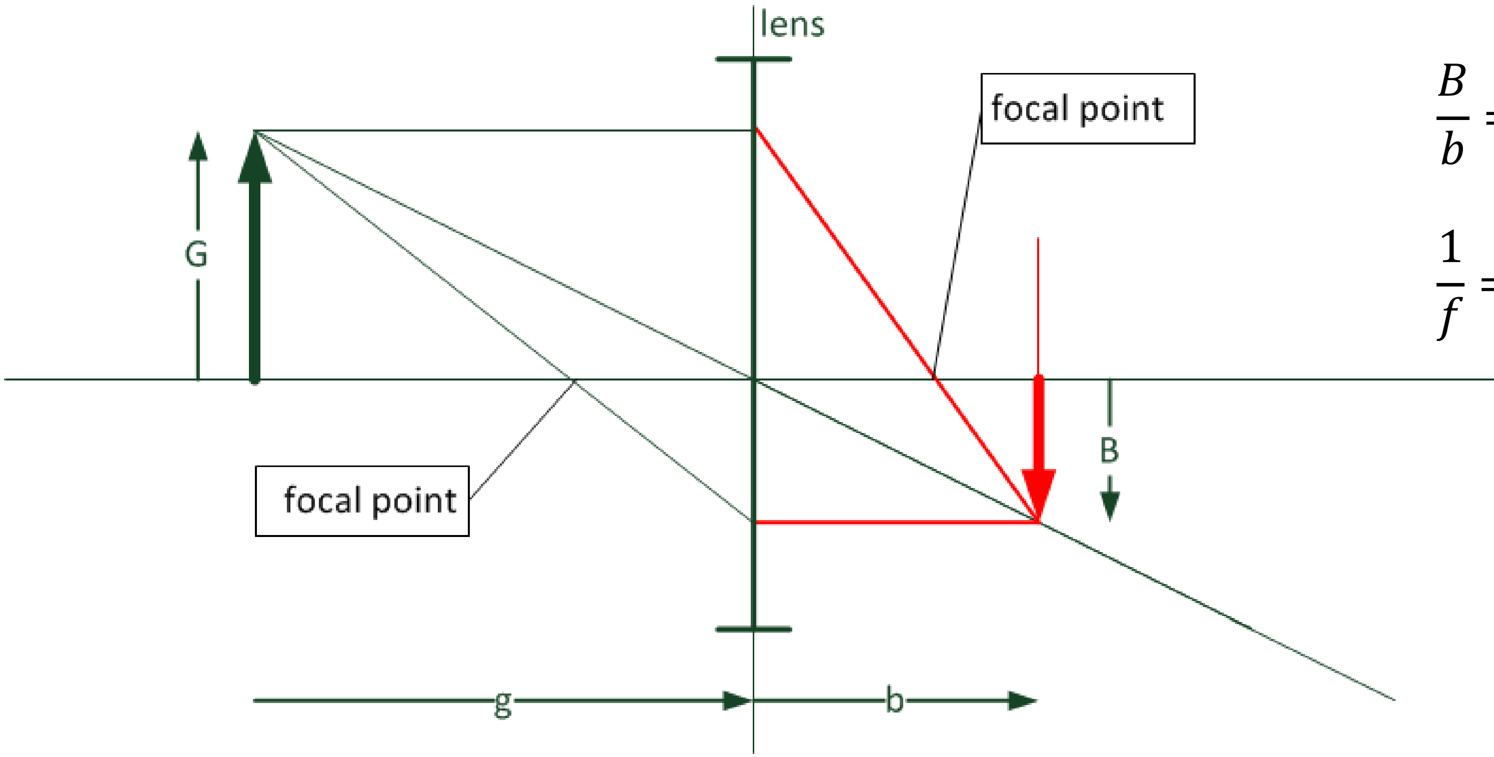
Relation with optics
Large sensor with same resolution and image section



Why do cameras with larger sensor deliver better results?

Fraunhofer IIS

Reminder: lens equations

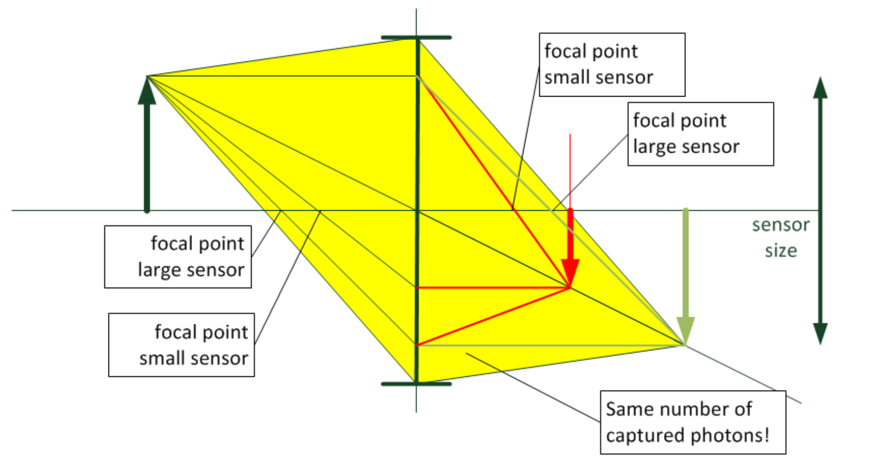


$$\frac{B}{b} = \frac{G}{g}$$

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b}$$

Scaling a camera

- Scale pixel extension by s
  - Area is scaled by s^2
- B needs to be scaled by s (same number of image pixels)
- $\frac{B}{b} = \frac{G}{g} \Rightarrow$ b needs to be scaled by s
- $\frac{1}{f} = \frac{1}{g} + \frac{1}{b \cdot s}$
- $\frac{1}{f} = \frac{1}{s} \cdot \left(\frac{1}{g} + \frac{1}{b}\right) + \frac{1 - \frac{1}{s}}{g}$
  - If $f \ll g$, f needs to be scaled by s



Fraunhofer IIS

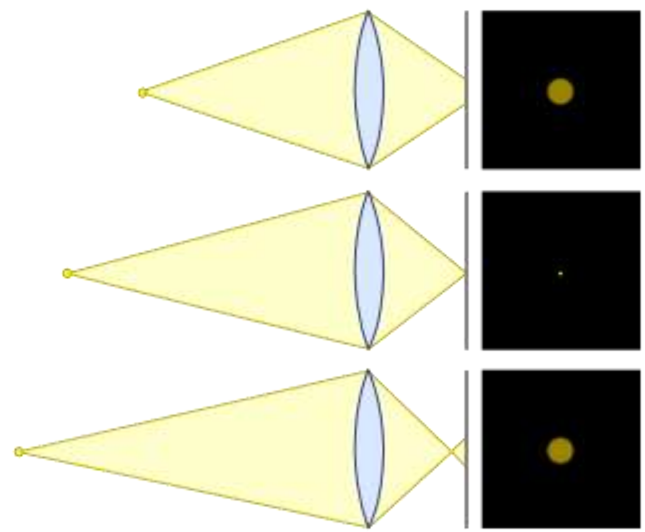
Scaling a camera
Impact on the optics

- F-number of lens: $k = \frac{f}{D}$ (D: aperture diameter)
  - Determines cost (and depth of field)
- Same F-number
  - D scales with s
  - Aperture area scales by s^2
  - Number of arriving photons is scaled by s^2 (for constant radiance)
- Larger pixels permit larger apertures
  - More photons per pixel
- Its only due to the combination with the lens, why larger pixels can deliver better SNR
- If D is kept constant, the number of photons hitting a pixel is constant!

Problematic for light field capture

Depth of field
Hyperfocal distance

- Finite object distance, such that objects with infinite distance have acceptable unsharpness
- $h = f + \frac{f^2}{k \cdot Z}$
  - f: focal length
  - k: f-number
  - Z: circle of confusion

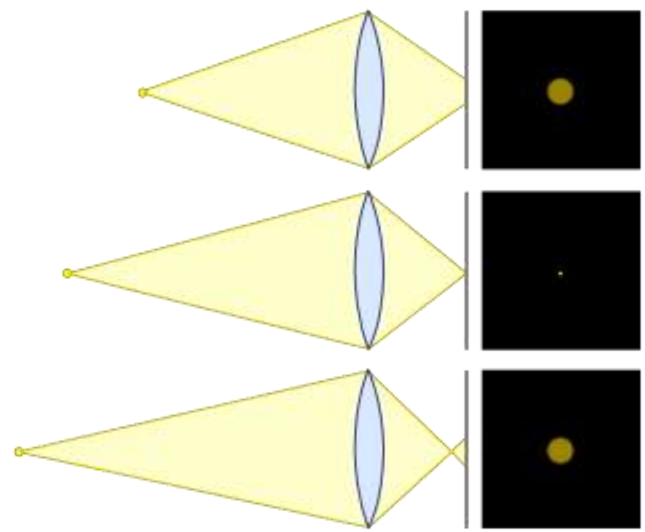


Depth of field
Hyperfocal distance

- Scaling the pixel size with s
  - Z can also be scaled with s

- $h = f + \frac{f^2}{k \cdot Z}$

- For constant k, h scales with s
- Larger pixel sizes increase h
  - Depth of field gets smaller



Depth of field
What happens for constant aperture size

- $h = f + \frac{f^2}{k \cdot Z}$
  - f: focal length
  - k: f-number
  - Z: circle of confusion
- When scaling the pixel size with s, Z can also be scaled with s

- Z scales with s
- For constant aperture size
  - $k = \frac{f}{D}$ scales with s
- In typical setups, $f \ll \frac{f^2}{k \cdot Z}$
- $h \approx \frac{f^2}{k \cdot Z}$ is approximately constant
- Depth of field is almost constant

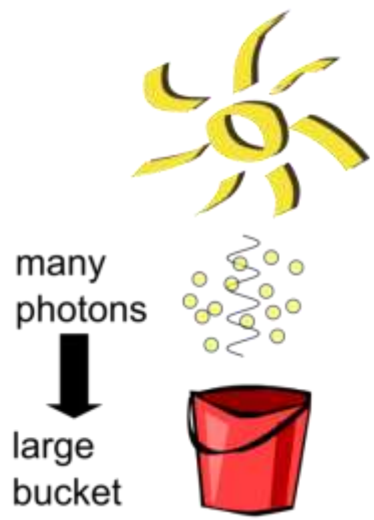
Conclusions on sensor size

- Cameras with larger sensor sizes can deliver the same image as a camera with smaller sensor
- Theoretically the same image quality
  - Requires equivalent sensor quality
  - Requires special tricks (see later)
- A camera with a larger sensor can operate with larger apertures sizes
  - Collects more photons in dark conditions
  - Not possible for small sensors
- Advantage of typical DSLR cameras
- Contradicts requirements of lightfields having a good depth of field

Large sensor sizes for lightfield capture

- Are large sensor sizes useless for lightfield capture?

- Not quite. They help for dynamic range.

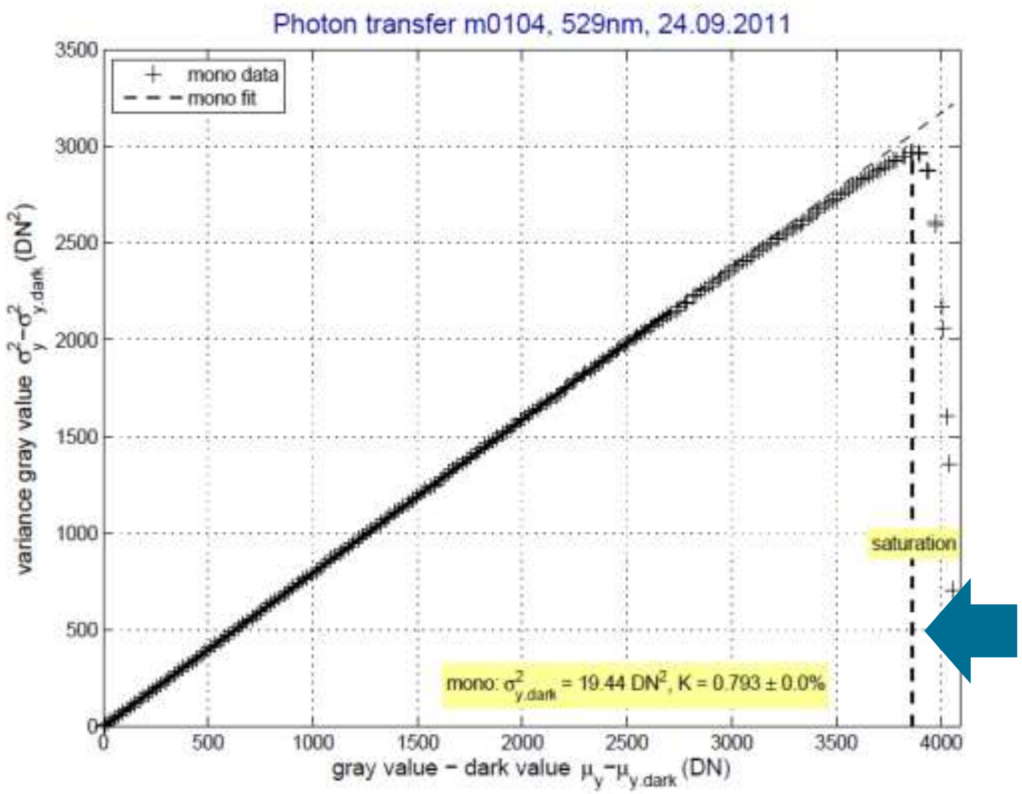


Fraunhofer
IIS

Definition of dynamic range
Sensitivity threshold

- Sensitivity threshold: SNR = 1
- $SNR \approx \frac{\eta \cdot \mu_p}{\sqrt{(\sigma_d^2 + \eta \cdot \mu_p) + \frac{\sigma_q^2}{K^2}}} = 1$
- $\eta \cdot \mu_p = \sqrt{(\sigma_d^2 + \eta \cdot \mu_p) + \frac{\sigma_q^2}{K^2}}$
- $\eta^2 \cdot \mu_p^2 - \eta \cdot \mu_p - \left(\sigma_d^2 + \frac{\sigma_q^2}{K^2}\right) = 0$
- $\mu_p = \frac{\eta \pm \sqrt{\eta^2 + 4\eta^2\left(\sigma_d^2 + \frac{\sigma_q^2}{K^2}\right)}}{2\eta^2}$
- $\mu_p = \frac{1}{2\eta} \cdot \left(1 + \sqrt{1 + 4\left(\sigma_d^2 + \frac{\sigma_q^2}{K^2}\right)}\right)$
- $\mu_p \approx \frac{1}{2\eta} \cdot (1 + 2\sigma_d)$

Definition of dynamic range
Saturation capacity



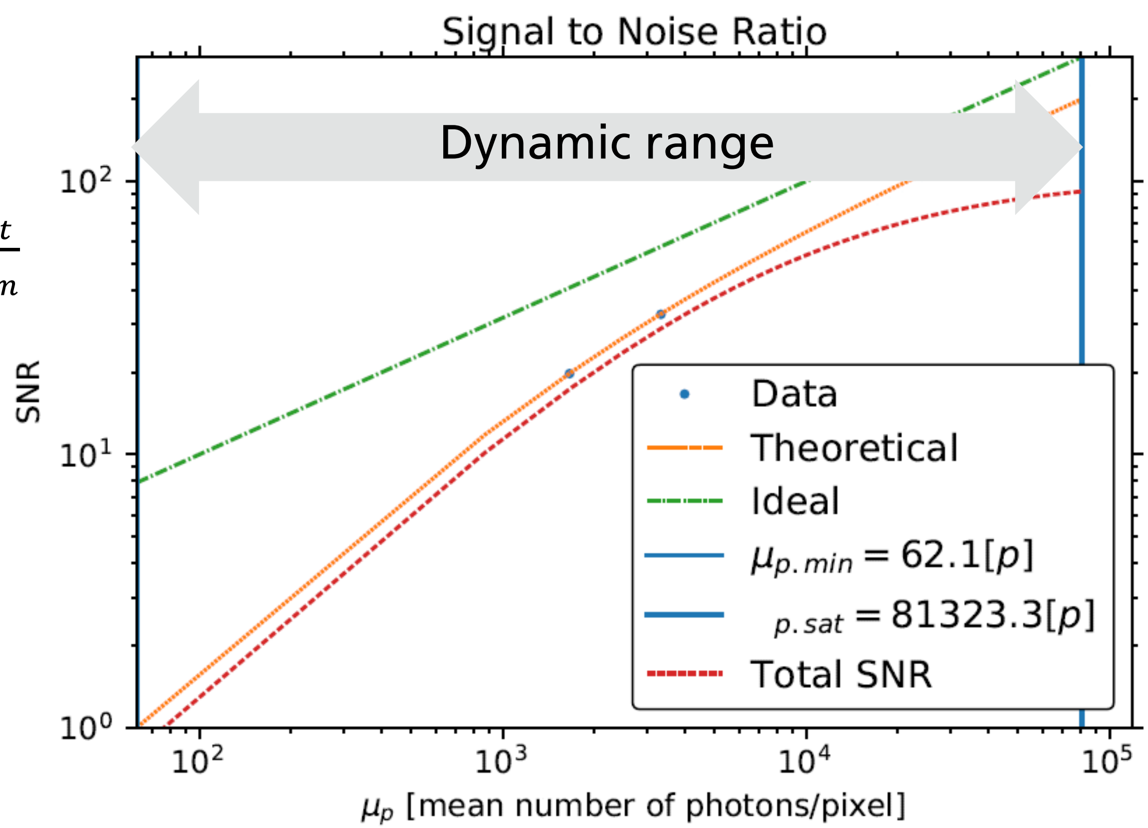
$\mu_{p,sat}$

Basler acA2500-14gm mit Aptina MT9P

Bernd Jähne, Was man über Bildsensoren wissen muss, um damit messen zu können, 58. Heidelberger Bildverarbeitungsforum

Model of a digital camera
Dynamic range

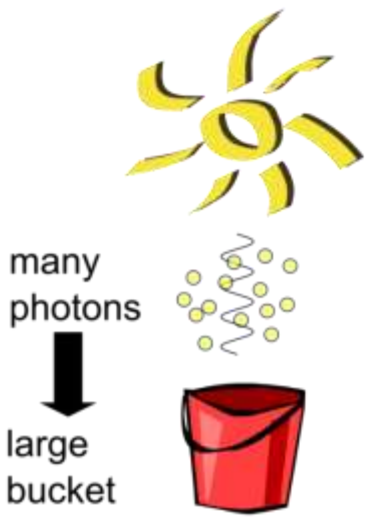
- $DR = \frac{\mu_{p,sat}}{\mu_{p,min}}$
- $DR_{dB} = 20lg\frac{\mu_{p,sat}}{\mu_{p,min}}$
- $DR_{stops} = \log_2\frac{\mu_{p,sat}}{\mu_{p,min}}$



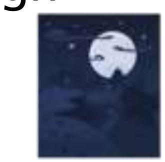
EMVA Standard 1288

Model of a digital camera
Building sensors with large dynamic range

- Upper bound determined by saturation capacity
  - Large full well capacity

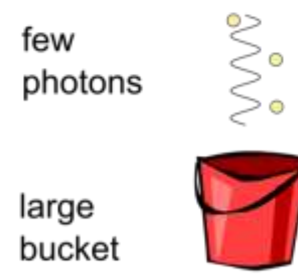


- $\mu_{p,sens} =$
  $$\frac{1}{2\eta} \cdot \left(1 + \sqrt{1 + 4\left(\sigma_d^2 + \frac{\sigma_q^2}{K^2}\right)}\right)$$
- Lower bound determined by noise and quantum efficiency η
  - Good electronics design

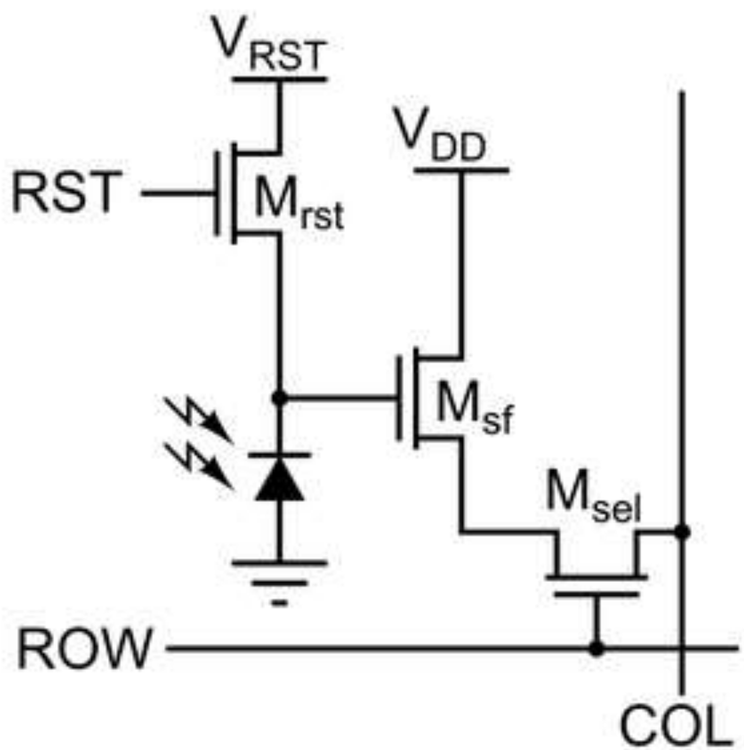


- Challenge:

  Collect few photons in large bucket



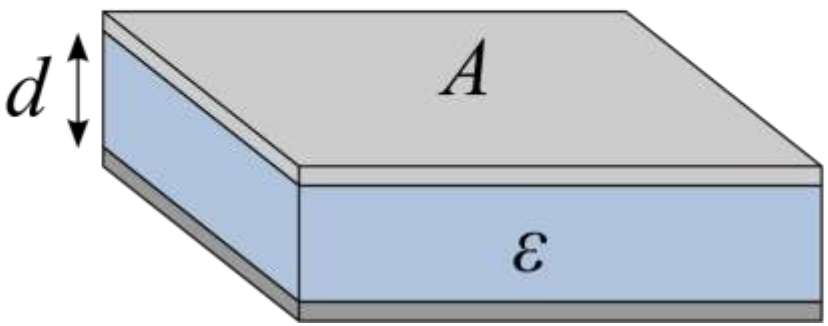
Model of a digital camera
Pixel cell (3 transistor)



http://commons.wikimedia.org/wiki/File:Aps_pd_pixel_schematic.svg

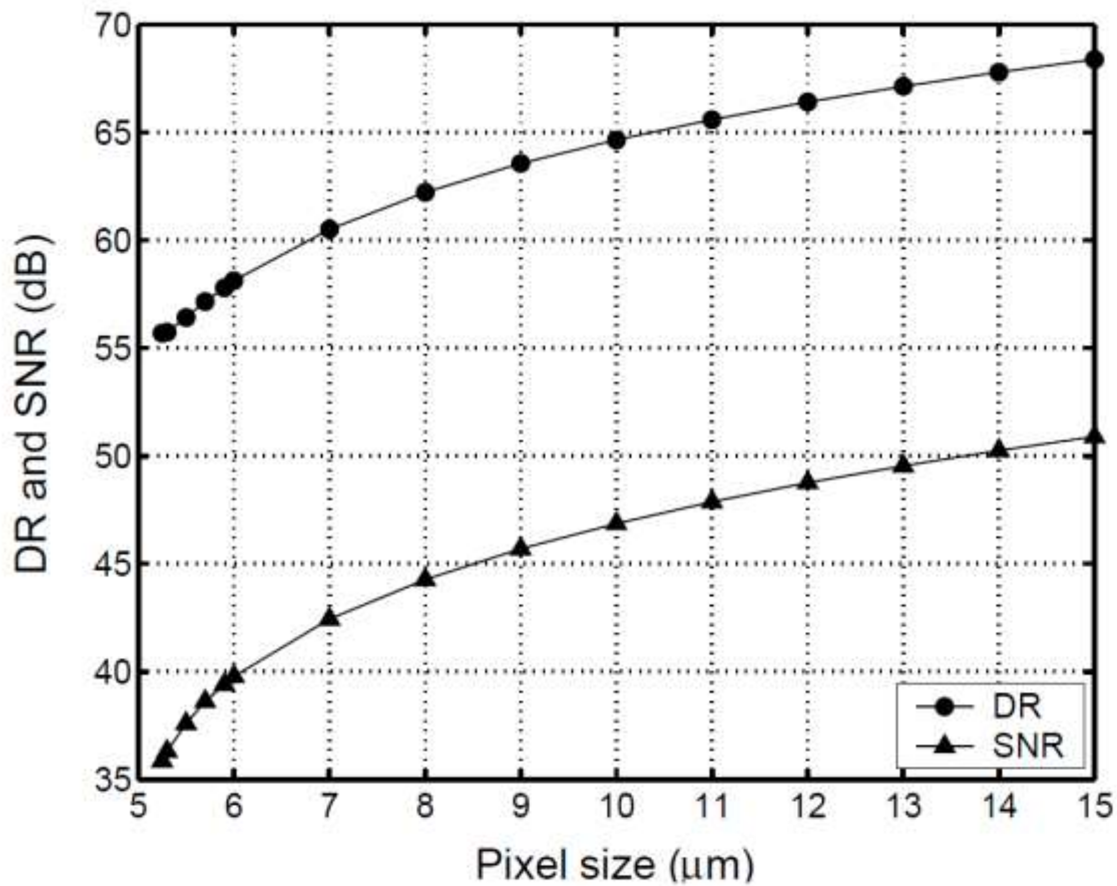
1. Reset photo diode by applying positive voltage to RST
   - Capacitor in photo diode will be charged
2. Release reset
   - Exposure takes place
   - Light makes photo diode conductive
   - Charges in capacitor are reduced
3. Row read-out
   - M_{sf} acts as a resistor

Capacitors



- $C = \frac{Q}{U}$
- Voltage U is limited due to small dimensions of semiconductor structures
- $C = \epsilon_0 \cdot \epsilon_r \cdot \frac{A}{d}$
  - ϵ_0: physical constant
  - ϵ_r: material dependent
  - d: Cannot be arbitrarily small
    - $E = \frac{U}{d}$ cannot exceed threshold
- A primarily determines capacity

Model of a digital camera
Building sensors with large dynamic range



Ting Chen, Peter Catrysse, Abbas El Gamal and Brian Wandell

How Small Should Pixel Size Be ?

Fraunhofer IIS

Image quality factors

- Large pixel sizes allow for high dynamic range
  - Large full well capacity
  - Requires very good read-out electronic to exploit (not further discussed in this presentation)
- Large pixel sizes **can** improve the SNR
  - Important for low-light conditions
  - Needs corresponding high quality lenses with larger apertures
  - Reduces depth of field
  - For same depth of field, SNR is almost the same than for smaller pixel sizes
  - Requires very good read-out electronic, since due to larger capacity, voltage change is smaller!!!

Professional Video Cameras
ARRI Alexa SXT

Sensor size	35 mm
Shutter	Rolling (+mechanical)
Dynamic range	14+ stops
Bit depth processing pipeline	16 bits
Frame rates	0.75-120 fps
Weight (body)	6.5 kg
Length	32.1 cm
Width	16.6 cm
Height	15.8 cm

Fraunhofer IIS

Professional Video Cameras
ARRI Alexa Mini

Sensor size	35 mm
Resolution	3422x2202
Pixel size	8.25 x 8.25 µm
Shutter	Rolling
Dynamic range	14+ stops
Frame rates	0.75 - 200
Weight (body)	2.3 kg
Length	18.5 cm
Width	12.5 cm
Height	14 cm

Professional Video Cameras
Sony F65

Sensor size	35 mm
Pixels	20M
Shutter	Rolling (+mechanical)
Dynamic range	14 stops
Bit depth processing pipeline	16 bit
Frame rates	Up to 120 fps
Weight (body)	5-6.5 kg

Camera arrays used for light field capture
Industrial cameras (Basler)

Sensor size	11.26 x 5.98 mm
Resolution	2046 x 1086
Pixel size	5.5 x 5.5 µm
Shutter	Global
Dynamic range	56.5 dB
Frame rates	Up to 50 fps
Weight (body)	90 g
Length	4.2 cm
Width	2.9 cm
Height	2.9 cm

Camera Arrays used for light field capture
Consumer cameras (GoPro)

Sensor size	5.37 x 4.04mm
Resolution	1920 x 1080
Pixel size	Ca. 3 µm
Shutter	Rolling
Dynamic range	
Frame rates	60 fps
Weight (body)	
Length	
Width	
Height	

Camera Arrays used for light field capture
Broadcast cameras (Blackmagic Studio)

Sensor size	13.06 x 7.344mm
Resolution	3840 x 2160
Pixel size	Ca. 3 µm
Shutter	Rolling/global
Dynamic range	11 f-stops
Frame rates	60 fps
Weight (body)	304g
Length	69.5mm
Width	82.5mm
Height	65.4mm

Fraunhofer IIS

Challenges in light field capture

- Larger pixel sizes lead to better dynamic range
  - For good performance in low light conditions (large SNR), large aperture sizes are required
  - Professional video cameras go for this option
- Light field algorithms require large depth of field
  - Aperture diameters need to be limited
  - Inherent challenge for low light captures