Basics for Enhanced Visualization: 3D/Data

Image formation

Rodrigo Cabral

Polytech Nice - Data Science

cabral@unice.fr

Source: by unknown, possibly Italian. The background shows Brunelleschi's Duomo, Florence. - Library of Congress[1] (image 500 of 601),[2], Public Domain, https://commons.wikimedia.org/w/index.php?curid=5597328

Outline

1. Introduction
2. Change of reference frame
3. Projection
   - Pinhole camera model and perspective
   - Cameras with lenses
4. Digitalization
5. Conclusions

These slides are partly inspired from Lionel Fillatre's slides.

What image formation is?

- To know how we get the 3D scene from images, we need to know how images are formed from 3D structure.
- The process of obtaining an image (2D) from a 3D scene is known as **image formation**.

Introduction

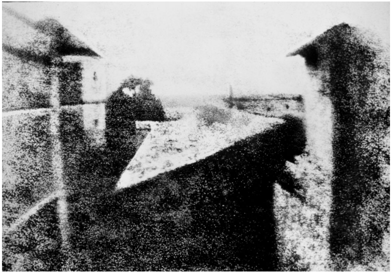
What image formation is?

- First imaging devices: *camera obscura*

- They were developed to prove the mathematical theory of perspective from Filippo Brunelleschi.
- The images could not be stored.

What image formation is?

- First imaging devices: *heliography*



- Predecessor of photography. They were invented by Joseph Nicéphore Nièpce - Picture is known as *Point de vue du Gras*.
- The images were stored on a pewter plate with a light sensitive material (bitumen of Judea).

Source: by Joseph Nicéphore Niépce - Rebecca A. Moss, Coordinator of Visual Resources and Digital Content Library, via email. College of Liberal Arts Office of Information Technology, University of Minnesota. http://www.dcl.umn.edu, Public Domain, https://commons.wikimedia.org/w/index.php?curid=107219

What image formation is?

Given a known 3D point in Cartesian coordinates in world frame $\mathbf{u}_w$:

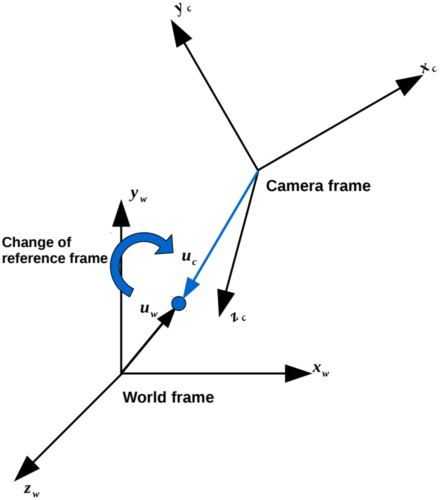
- What the camera sees in 3D? $\Longrightarrow$ Change from world frame to the camera frame.
  - Here we call a **camera** any imaging device.
- How the camera projects its 3D view to 2D? $\Longrightarrow$ Perspective projection and camera internal parameters.
- How do the camera transforms an analog image into a digital image in pixels? $\Longrightarrow$ Digitalization.

Change the reference from the world frame to the camera frame

- Camera is an object with its own position, translation vector $\mathbf{t}_{cw}$, and orientation, rotation matrix $\mathbf{R}_{cw}$, with respect to the world frame.
- Camera viewing direction points to negative part of its own z_c axis.
- The world frame position and orientation is also given. Normally it is fixed by some object in the image (the checker board reference in the OpenCV example).

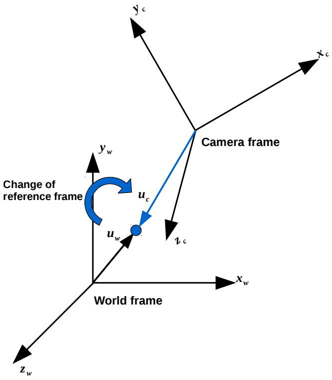
Change of reference frame

Change the reference from the world frame to the camera frame



Change of reference frame

Change the reference from the world frame to the camera frame



To get vector $\mathbf{u}_c$:

- Translate points so that $\mathbf{t}_{cw}$ becomes the origin.
- Rotate points so that $\mathbf{x}_c$ and $\mathbf{y}_c$ becomes aligned with $\mathbf{x}$ and $\mathbf{y}$.

We often call $\mathbf{R}_{wc} = \mathbf{R}_{cw}^{-1}$. Thus,

$$\mathbf{u}_c = \mathbf{R}_{wc}(\mathbf{u}_w - \mathbf{t}_{cw})$$

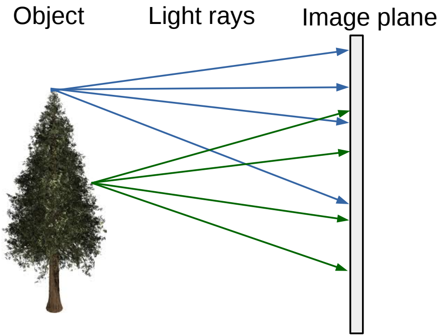
In homogeneous coordinates

$$\begin{bmatrix} \mathbf{u}_c \\ 1 \end{bmatrix} = \mathbf{K}_{\text{wc}} \begin{bmatrix} \mathbf{u}_w \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{wc} & -\mathbf{R}_{wc}\mathbf{t}_{cw} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_w \\ 1 \end{bmatrix}$$

Projection

Pinhole camera model

Imaging device without any barrier

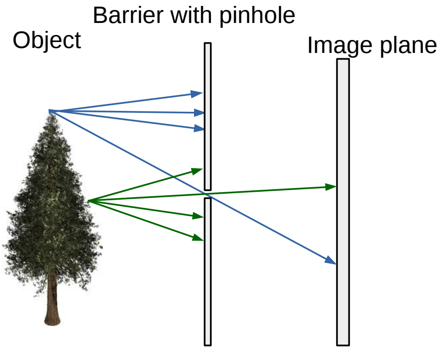


- Rays reflected from the same point are all scattered through the image plane. $\Longrightarrow$ The image is blurry.

Source for the pine tree:https://pixabay.com/en/tree-evergreen-isolated-pine-1702027/

Pinhole camera model

Imaging device with a barrier

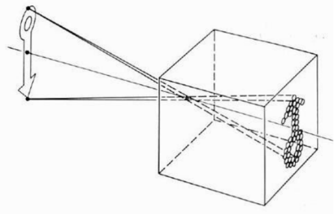


- Pinhole size is called **aperture**.
- Pinhole ideally lets only one ray pass from each point.
- Image is not blurry.

Projection

Pinhole camera model

Imaging device with a barrier



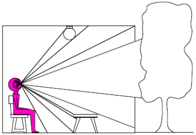
- Image plane capture the pencil of light rays.
- The pinhole is called **center of projection** or **focal point**.
- Barrier plane is called **focal plane**.
- The image formed in image plane has reversed coordinates.

Projection

Pinhole camera model

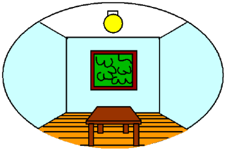
Projection on the image plane = dimensionality reduction

3D world



Point of observation

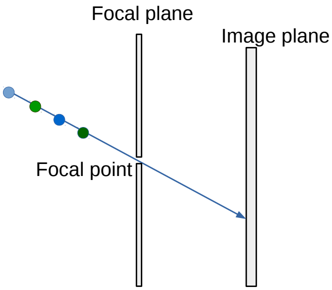
2D image



Figures © Stephen E. Palmer, 2002

- We lose the angles.
- We lose the distances.

Pinhole camera model



Properties of the projection:

- All points in a ray give one point in image.
- Points are still points. Points on focal plane are undefined.
- All lines remain lines except for the line passing through the pinhole.
- All planes remain planes (or half-planes) except planes which contain the focal point.

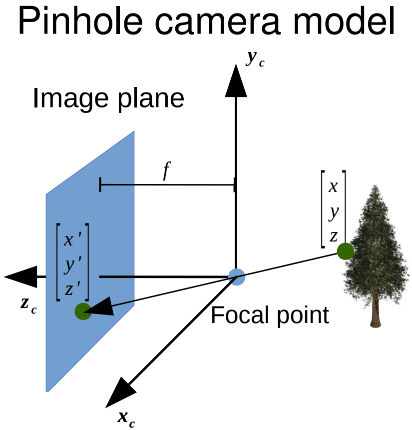
Projection

Pinhole camera model

- Parallel lines converge to a point.
- Each direction has a vanishing point.
- Only exceptions are lines parallel to image plane which are parallel.

Source: https://pixabay.com/en/railway-rails-track-railroad-731340/

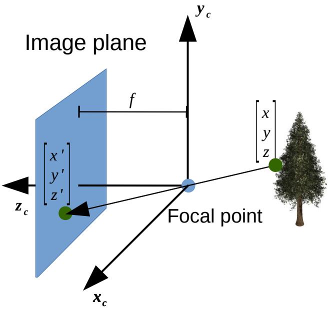
Projection



- Focal point is at the origin of camera frame.
- Image plane is parallel to $x_c y_c$ plane and at $z_c = f$.
- Axis z_c is called **optical axis**.
- f is called **focal distance** or **focal length**.

Projection

Pinhole camera model



- $\begin{bmatrix} \alpha x \\ \alpha y \\ \alpha z \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ \mathbf{f} \end{bmatrix}$ - light ray is a line passing through the origin.
- Therefore $\alpha = f/z$ and $\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} (f/z)x \\ (f/z)y \\ f \end{bmatrix}$.

Pinhole camera model

- From the 3D point in homogeneous coordinates we map to a 2D point in the image plane also in homogeneous coordinates:

$$\begin{bmatrix} x'_h \\ y'_h \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- Switching to Cartesian coordinates we get the right coordinates:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x'_h/w \\ y'_h/w \end{bmatrix} = \begin{bmatrix} (f/z)x \\ (f/z)y \end{bmatrix}$$

Pinhole camera model

Previous transformation is normally described as a composition:

$$\mathbf{u}'_h = \mathbf{K}_f \mathbf{\Pi}_0 \mathbf{u}_h$$

where

$$\mathbf{\Pi}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$ is called the **canonical projection matrix**

and

$$\mathbf{K}_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ is a scaling in 2D homogeneous coordinates.

Projection

What are the transformations so far?

World to camera transformation

$\Downarrow$

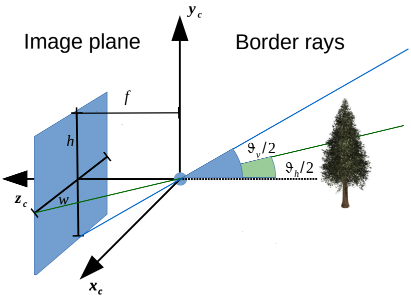
Canonical projection

$\Downarrow$

Scaling

$$\mathbf{u}'_h \quad = \quad \mathbf{K}_f \, \boldsymbol{\Pi}_0 \, \mathbf{K}_{\mathrm{wc}} \, \mathbf{u}_w$$

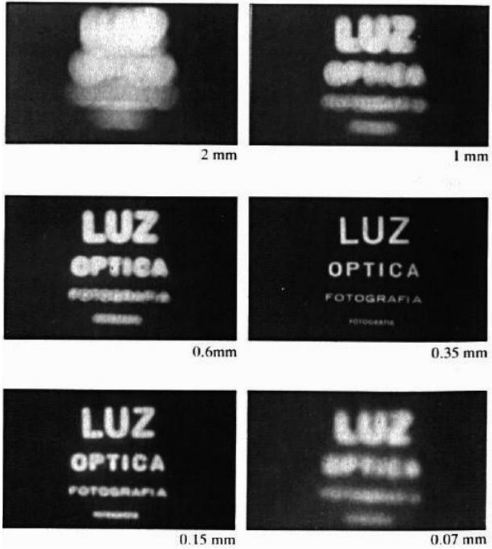
Field of view for finite image plane



- Image plane has finite width w and height h.
- **Field of view** angles: vertical $\theta_v = 2\tan^{-1}\left(\frac{h/2}{f}\right)$
  horizontal $\theta_h = 2\tan^{-1}\left(\frac{w/2}{f}\right)$
- Choice of f: trade off between resolution and field of view.

Cameras with lenses

Issues with pinhole cameras



Large aperture:
too blurry but right intensity.

Small aperture:
relatively clear but faint.

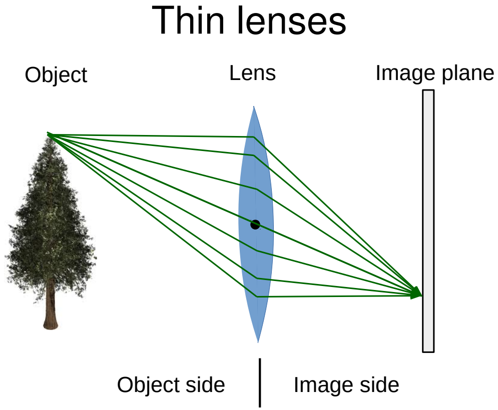
Very small aperture:
too blurry and still faint. Diffraction!

Cameras with lenses

How to increase aperture without increasing blur?

Solution: thin lenses

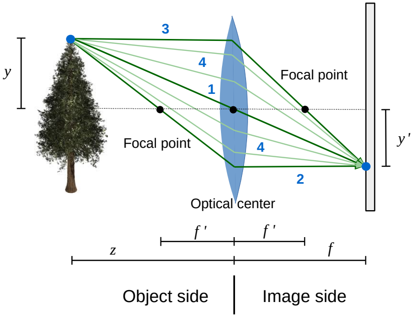
Cameras with lenses



- Focus rays from an object point in image plane without small pinhole.

Cameras with lenses

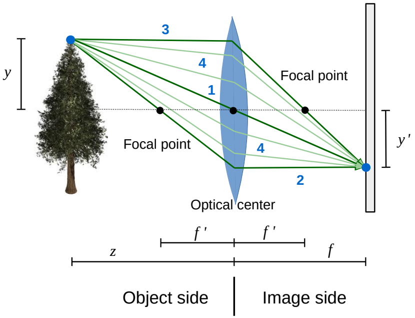
Convergent thin lenses: properties



1. Rays passing through optical center are not deflected.
2. Rays passing through focal point in object side become parallel to optical axis.
3. Rays which are parallel to optical axis in the object side pass through the focal point in image side.
4. Rays from the same point in object side will converge to the same point point in image side.

Cameras with lenses

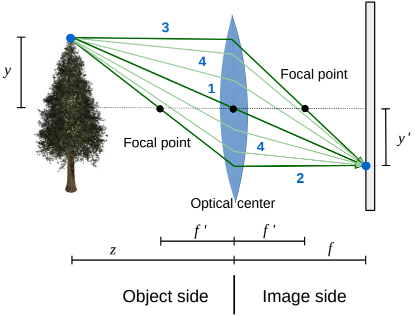
Convergent thin lenses



- Focal point here means a different thing than in a pinhole camera.
- For a fixed image plane distance f there is a distance z where objects are perfectly focused
- The focused distance z can be retrieved from the intersection of lines 1and 3.

Cameras with lenses

Convergent thin lenses

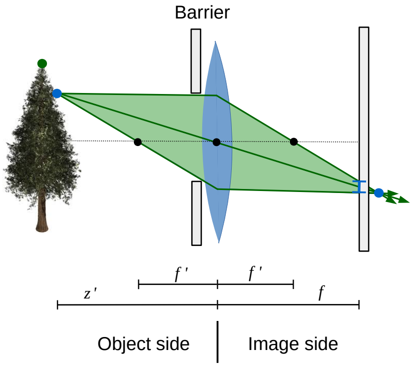


- Perfectly focused point position is given by

$$\frac{1}{z} + \frac{1}{f} = \frac{1}{f'}$$

Cameras with lenses

Convergent thin lenses



- If image plane is fixed, points in other position are not focused
- If a circular barrier is present, a point generates a circular blurry pattern, which is called **circle of confusion**.

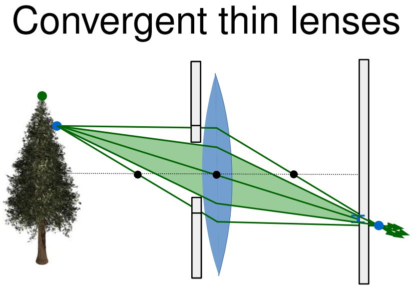
Cameras with lenses

Convergent thin lenses

- For a fixed imaging device, the range of distances where the circle of confusion is of acceptable size is called **depth of field**.

Source: https://upload.wikimedia.org/wikipedia/commons/1/12/Fence_Illustrating_Narrow_Depth_Of_Field_In_Photography._Shepperton_UK.jpg

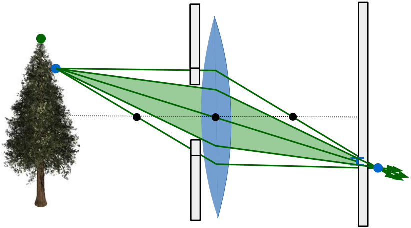
Cameras with lenses



- To increase the depth of field we can reduce the aperture.
- But if the aperture is reduced too much we get a faint image (like a pinhole camera).

Cameras with lenses

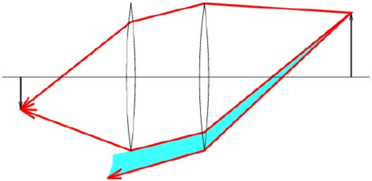
Convergent thin lenses



- Central point of circle of confusion is given by ray passing through optical center.
- Therefore, for the central point, **we can use the pinhole camera model**.

Cameras with lenses

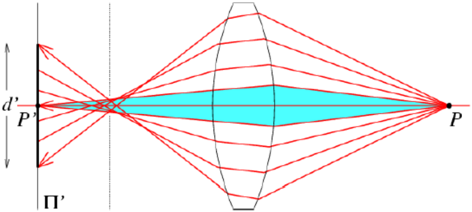
Non ideal lenses: vignetting



- Light intensity is blocked in peripheral areas
- Objects and points in peripheral areas are difficult to track.

Cameras with lenses

Non ideal lenses: spherical aberration



- Spherical lenses do not focus perfectly.
- Rays passing through lens edge are focused closer.
- Increase in circle of confusion, even in the focused position.

Cameras with lenses

Non ideal lenses: radial distortion

Correct Barrel distortion Pin-cushion distortion

- Image space is distorted radially.
- Distances between points decrease radially in barrel distortion.
- Distances between points increase radially in pin-cushion distortion.
- From $[x' \quad y']^T$ we can get the distorted point $[x_d \quad y_d]^T$ with a polynomial mapping:

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + \kappa_1 r^2 + \kappa_2 r^4) \begin{bmatrix} x' \\ y' \end{bmatrix}$$

where $r^2 = (x')^2 + (y')^2$ and κ_1 and κ_2 are fixed coefficients.

Source: https://www.cvl.isy.liu.se/education/graduate/geometry-for-computer-vision-2014/geometry2014/lecture4.pdf

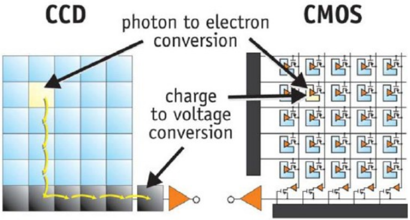
Digitalization

From analog to digital

- Digital cameras replace film with an array of photon sensitive diodes $\Longrightarrow$ transformation of light intensities in space into pixels.
- Diodes collect photons in a given surface and transforms it into electrons.
- The electrons in diodes are collected in two different ways:
  - Coupled charged device (CCD).
  - Complementary metal oxide semiconductor (CMOS).

Digitalization

From analog to digital (bottom to top)



CCDs move photogenerated charge from pixel to pixel and convert it to voltage at an output node. CMOS imagers convert charge to voltage inside each pixel.

- **CCD** transports the charges across the chip and reads at the corner of the array. An analog to digital converter (ADC) then transforms the measured amount of charges in a digital value (binary value).
- **CMOS** uses several transistors directly attached to the diodes to amplify and transform directly the amount of charges in a binary value. It is not a serial architecture and it does not require an additional ADC.

Digitalization

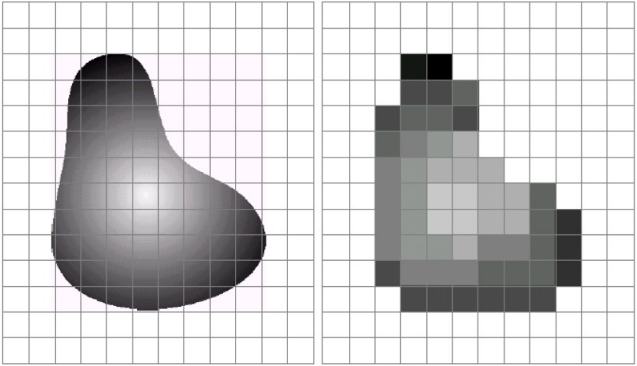
From analog to digital

Looking at upper level we have two main operations:

- Sampling
- Quantization

Digitalization

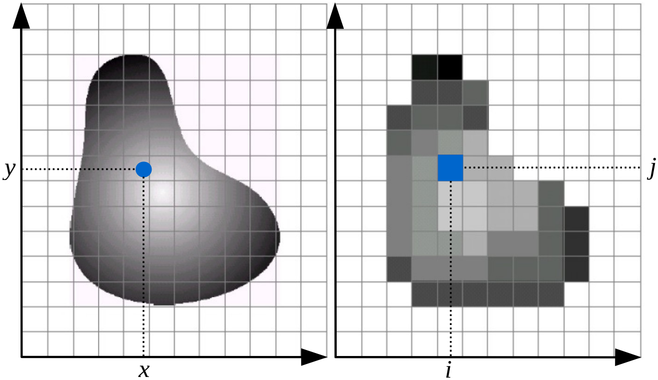
From analog to digital: sampling



- 2D continuous space $\Longrightarrow$ 2D discretized space.
- Key parameters: pixels per meter in x: s_x pixels per meter in y: s_y

Digitalization

From analog to digital: sampling



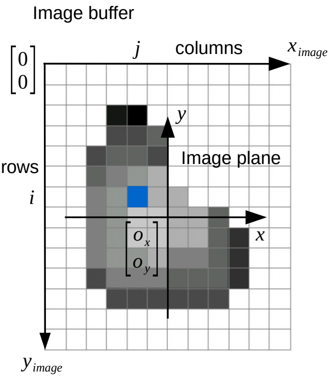
- The true relation between continuous and sampled positions is

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} \mathrm{nint}\,(s_x x) \\ \mathrm{nint}\,(s_y y) \end{bmatrix}, \quad \mathrm{nint}(\cdot) \text{ is rounding.}$$

- In practice, we keep linear approx.: $\begin{bmatrix} x_s \\ y_s \end{bmatrix} \approx \begin{bmatrix} s_x x \\ s_y y \end{bmatrix}$

Digitalization

From analog to digital: sampling and image buffer



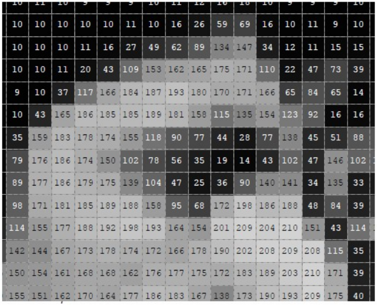
- This amounts to the following overall transformation:

$$\begin{bmatrix} x_{\mathrm{im}} \\ y_{\mathrm{im}} \end{bmatrix} = \begin{bmatrix} s_x x + o_x \\ s_y y + o_y \end{bmatrix}$$

where s_y can take into account the change of sign in y.

Digitalization

From analog to digital: quantization



10	10	10	10	10	11	10	16	26	59	69	16	10	11	9	10
10	10	10	11	16	27	49	62	89	134	147	34	12	11	15	15
10	10	11	20	43	109	153	162	165	175	171	110	22	47	73	39
9	10	37	117	166	184	187	193	180	170	171	166	65	84	65	14
10	43	165	186	185	185	189	181	158	115	135	154	123	92	16	16
35	159	183	178	174	155	118	90	77	44	28	77	138	45	51	88
79	176	186	174	150	102	78	56	35	19	14	43	102	47	146	102
89	177	186	179	175	139	104	47	25	36	90	140	141	34	135	33
98	171	181	185	189	188	158	95	68	172	198	186	188	48	84	39
114	155	177	188	192	198	193	164	154	201	209	204	210	151	43	114
142	144	167	173	178	174	172	166	178	190	202	208	209	208	115	35
150	154	161	168	168	162	176	177	175	172	183	189	203	210	171	39
155	151	162	170	164	177	186	183	167	138	173	190	193	209	175	40

- Discretize the amplitudes.
- Standard: 256 integer values per color (RGB)$\Longrightarrow$ 8 bits.
- 0 = color turned off, 255 = color completely turned on.
- High dynamic range (HDR): new encoding, floating point with 32 bits.

Digitalization

From analog to digital: overall effect

- For position of point sources quantization is not important.
- But quantization may have great effect on previous image processing.
- Overall approximate effect of digitalization is a scale and a translation.
- Transformation matrix in homogeneous coordinates is

$$\mathbf{K}_s = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

where s_θ is an additional parameter for non rectangular pixels.

Conclusions

What is the overall transformation?

World to camera transformation

$\Downarrow$

Canonical projection

$\Downarrow$

Scaling

$\Downarrow$

Digitalization

$$\mathbf{u}'_h = \mathbf{K}_s \, \mathbf{K}_f \, \mathbf{\Pi}_0 \, \mathbf{K}_{\mathrm{wc}} \, \mathbf{u}_w$$

Conclusions

$$\mathbf{u}'_h \quad = \quad \mathbf{K}_s\, \mathbf{K}_f\, \mathbf{\Pi}_0\, \mathbf{K}_{wc}\, \mathbf{u}_w$$

- u_w is in meters while u'_h is a pixel number in homogeneous coordinates.
- Remember that third coordinate of u'_h might not be 1 due to the projection.
- For radial distortion model, non linear function should be applied just before $\mathbf{K}_f$ and in Cartesian coordinates.
- In practice all parameters of these matrices are unknown. Retrieving them is called **camera calibration**. Next class.