3D Sensors

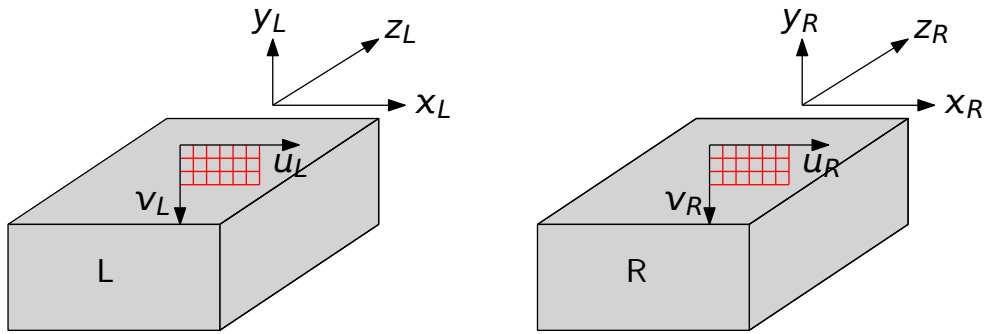
(Source: https://www.google.com/selfdrivingcar/)

Slide credit to Radu Horaud, http://perception.inrialpes.fr

Structured Light Imaging

Pinhole camera model & binocular stereo

To understand the principles of structured light imaging, we introduce a minimal stereo system (calibrated and rectified) first.



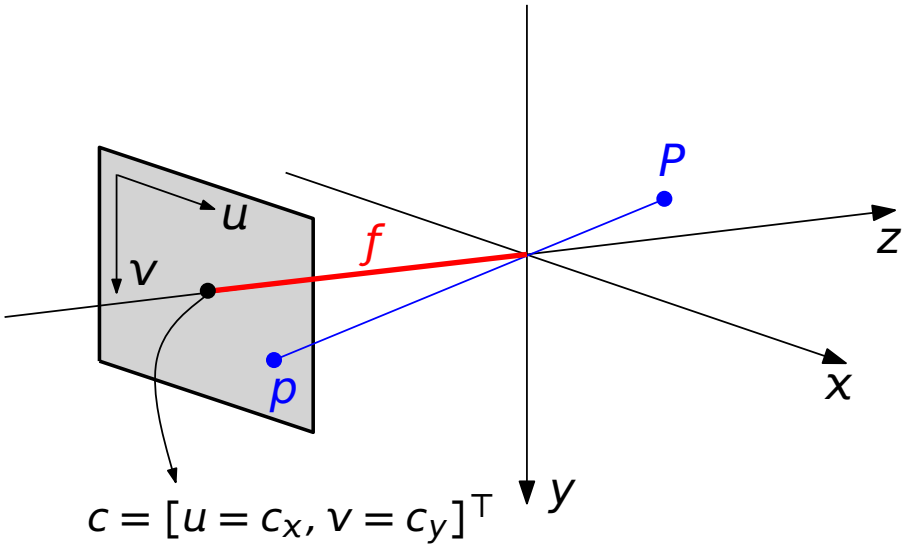
We adhere to the following convention:

- Left camera (L): reference camera
- Right camera (R): target camera

Structured Light Imaging

Pinhole camera model & binocular stereo

Pinhole camera model



We see that

$$u - c_x = f\frac{x}{z}$$
$$v - c_y = f\frac{y}{z}$$

and thus

$$z[u, v, 1]^\top = \mathbf{K}[x, y, z]^\top$$

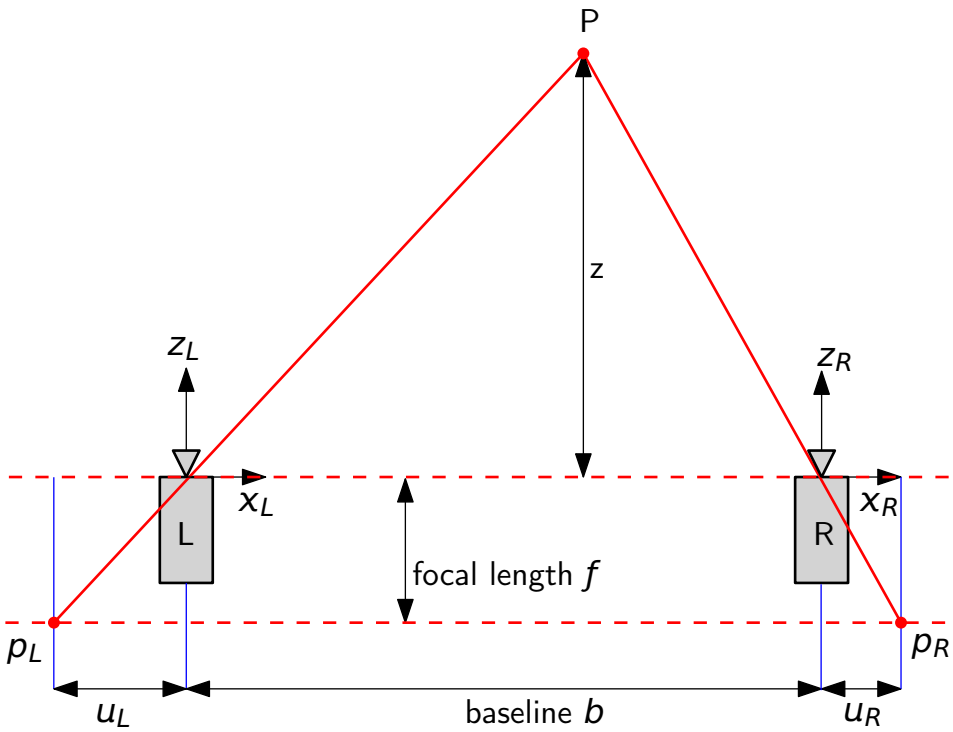
with $\mathbf{K}$ defined as

$$\mathbf{K} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_x \\ 0 & 0 & 1 \end{bmatrix}$$

Structured Light Imaging

Pinhole camera model & binocular stereo – Disparity vs. depth

Lets consider a point $P = [x, y, z]^\top$



The point is projected to p_L and p_R with $p_L = [u_L, v_L]^\top$ and $u_R = [u_L - d, v_L]^\top$. We denote $d = u_L - u_R$ as the disparity.

Structured Light Imaging

Pinhole camera model & binocular stereo – Disparity vs. depth

The relationship between disparity d, the z-coordinate of P (depth) and the focal length f is (via comparing similar triangles)

$$z = \frac{b|f|}{d}$$

Once we have a couple of conjugate pixel p_L and p_R and depth z, we can compute

$$[x, y, z]^\top = \mathbf{K}_L^{-1}[u_L, v_L, 1]^\top$$

with $\mathbf{K}_L$ being the intrinsic parameter matrix of a camera (here: for camera L).

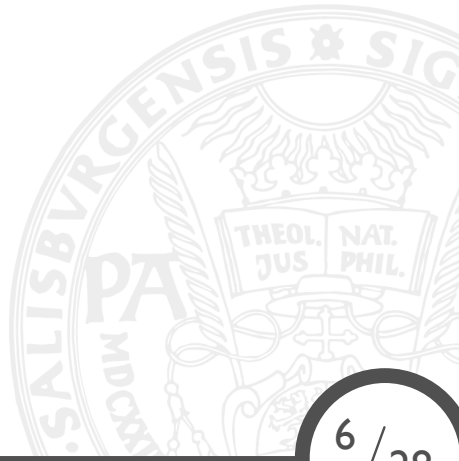
Structured Light Imaging

From cameras to projectors

Detecting the conjugate pixel pairs is known as the correspondence problem in stereo vision (which is typically the tricky part).

How do we get p_L and p_R?

- From light being reflected off P towards the cameras L and R
- But is this really the relevant point?



Structured Light Imaging

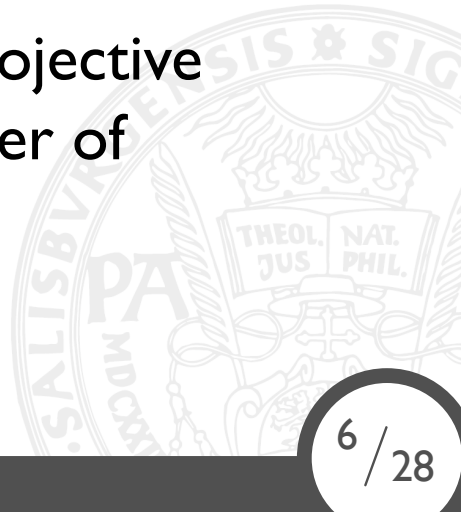
From cameras to projectors

Detecting the conjugate pixel pairs is known as the correspondence problem in stereo vision (which is typically the tricky part).

How do we get p_L and p_R?

- From light being reflected off P towards the cameras L and R
- But is this really the relevant point?

In fact, important is only the triangle geometry! In projective geometry, points are equivalent to rays exiting a center of projection.



Structured Light Imaging

Light coding system

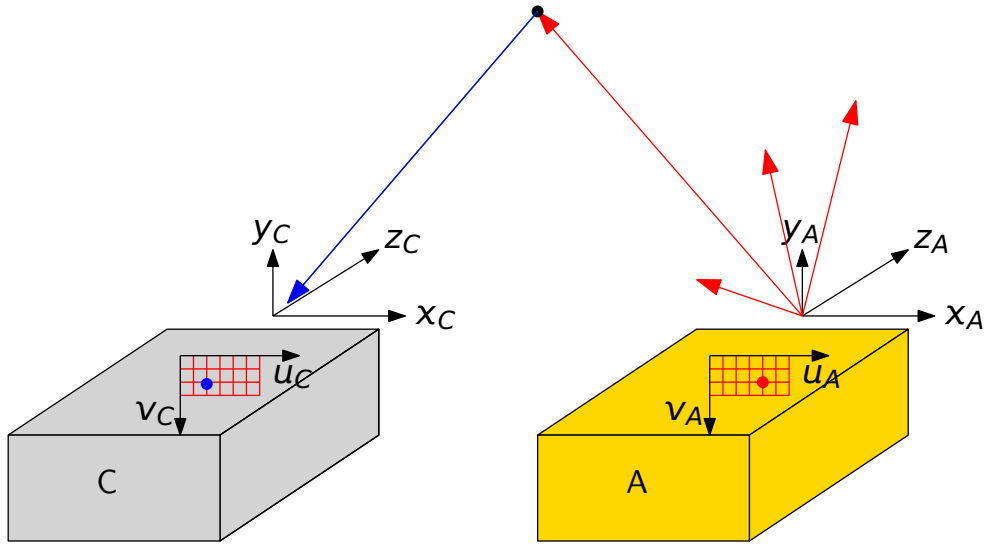
Definition (Light coding system)

A stereo system, in which one of the two cameras is replaced by a projector is called a *light coding system*.

Structured Light Imaging

From cameras to projectors

Camera R is replaced by the projector A



If P is not occluded, it projects the light, emitted by A, to pixel p_C of camera C.

As before, p_A (point to be projected) and p_C are conjugate points.

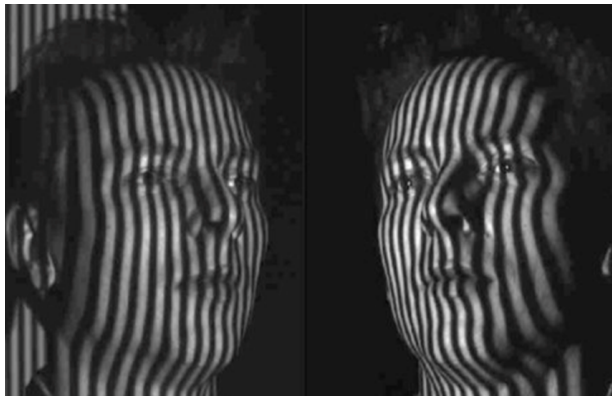
Structured Light Imaging

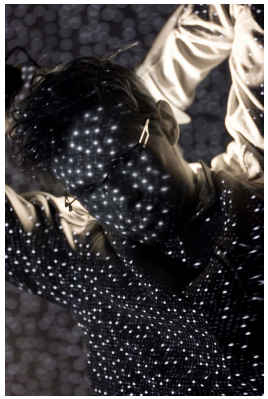
From cameras to projectors

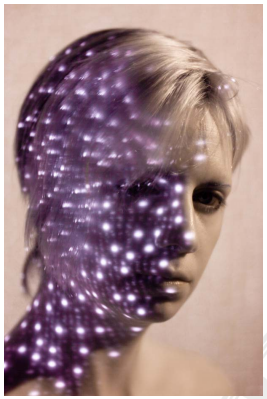
Example: Consider a straight wall with uniform color

- Projector A "colors", with its radiant power, the scene point P
- Its "easy" to identify p_C from its neighbors, due to color of P

In general, a suitable light pattern needs to be adopted!







(Pictures from Wikipedia and artist Audrey Penven)

Structured Light Imaging

Microsoft's Kinect™

- Light-coded range camera
- 30 frames per second (FPS) with VGA 640×480 resolution
- Equipped with color video camera and microphones
- Based on the Primesensor™ chip (also used by Asus X-tion)
- Uses Infrared (IR) light-coded patterns

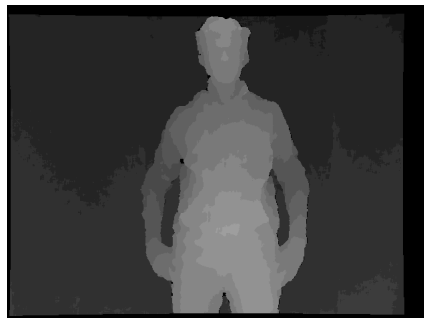
Image source: http://www.futurepicture.org

Structured Light Imaging

Microsoft's Kinect™– Available data (at 30 [FPS])

- IR Image I_K: defined on lattice Λ_K; values in $[0, 1]$
- Disparity map D_K: defined on lattice Λ_K; values in $[d_{min}, d_{max}]$
- Depth map Z_K: defined on lattice Λ_K; values in $[z_{min}, z_{max}]$ with

$$z_{min} = \frac{bf}{d_{max}} \quad \text{and} \quad z_{max} = \frac{bf}{d_{min}}$$


From *left* to *right*: I_K, D_K and Z_K.

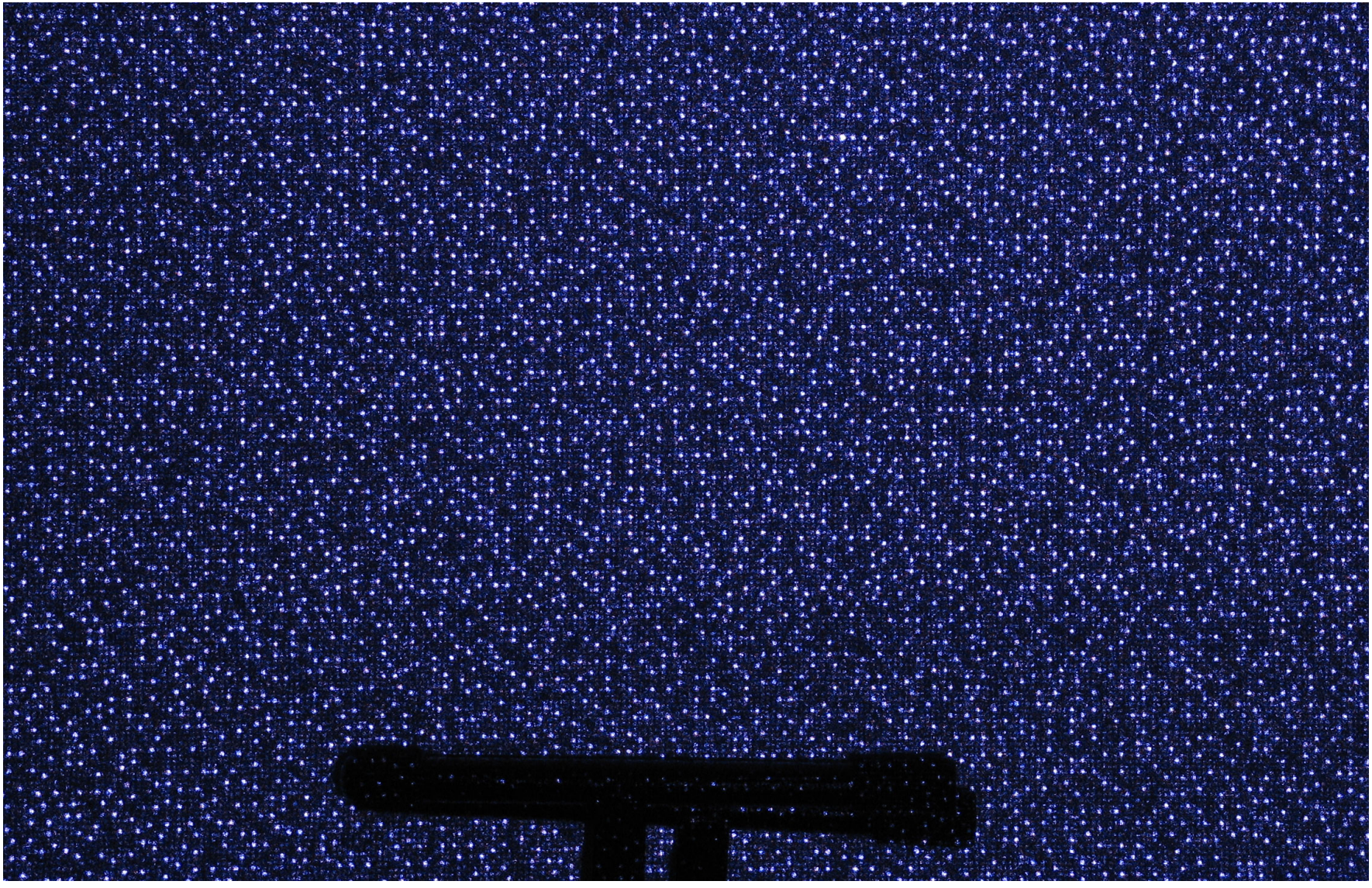


Image source: http://www.futurepicture.org

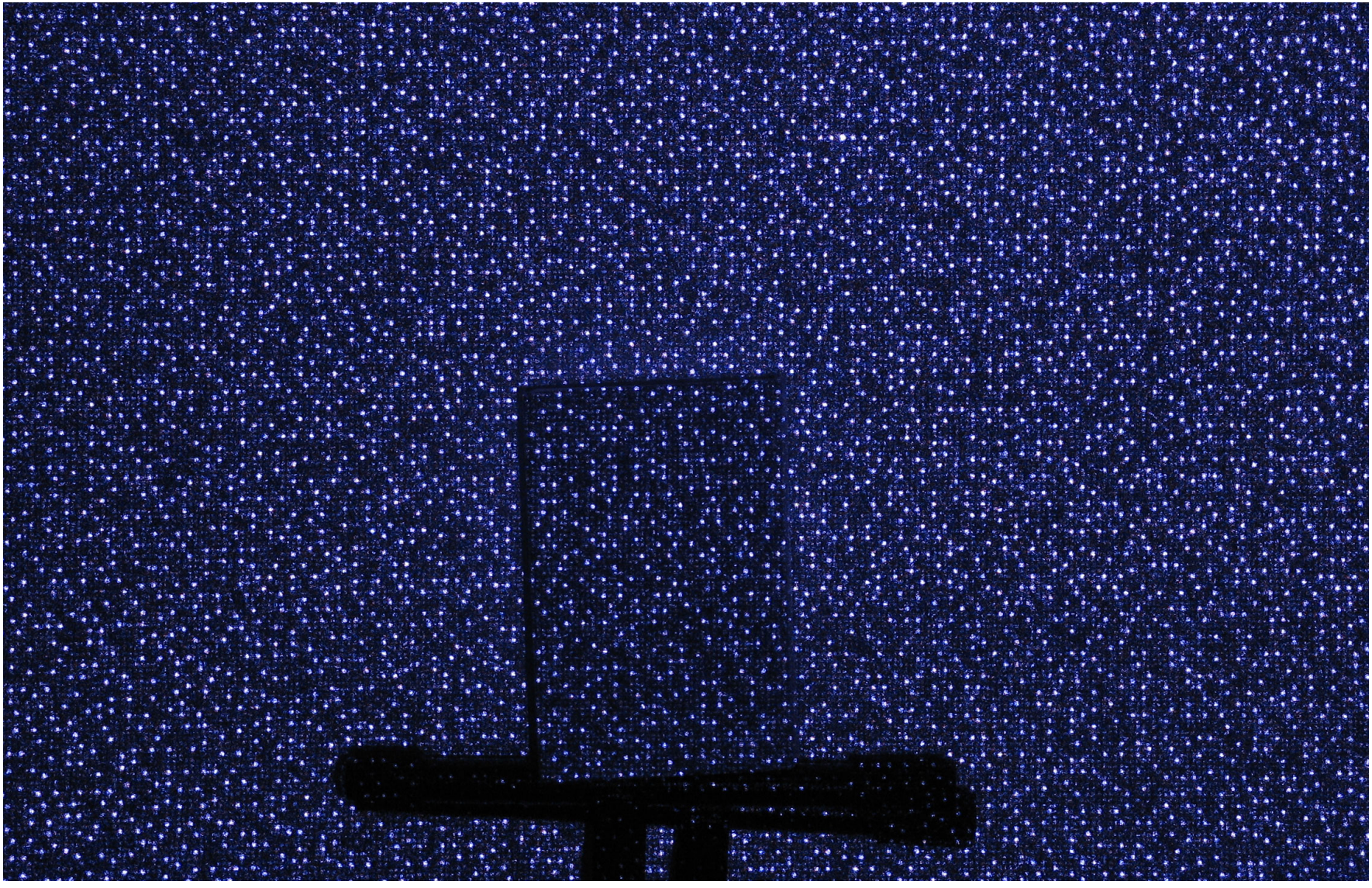


Image source: http://www.futurepicture.org

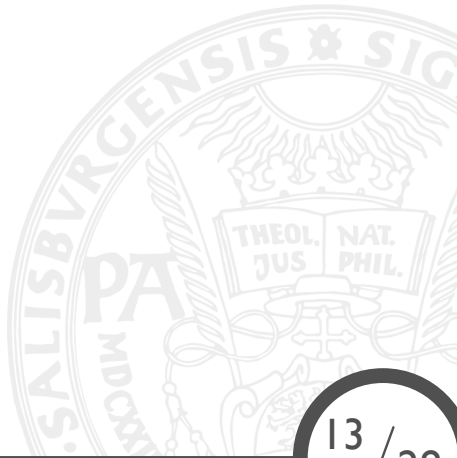
Structured Light Imaging

Microsoft's Kinect™

Some more Kintect™ internals:

- Baseline b: 75 [mm]
- Focal length f: $\sim$ 585.6 [px]

Hence, we get (in pixel): $(d_{min}, d_{max}) = (2, 88)$

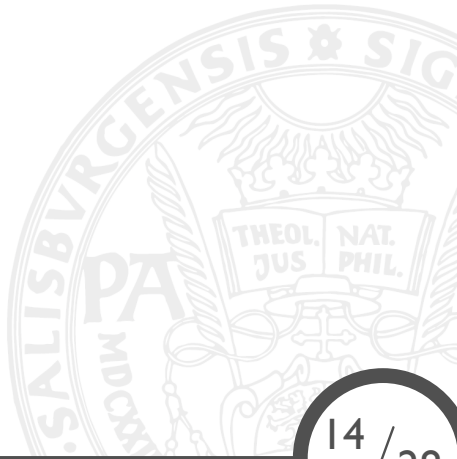


Structured Light Imaging

Light coding principles – What is the goal?

Goal of pattern design: decodable in the presence of non-idealities

- Assume that the projected pattern has $N_R \times N_C$ pixel p_A^i
- Active triangulation requires one codeword per pixel
- The more codewords are different $\rightarrow$ better robustness



Structured Light Imaging

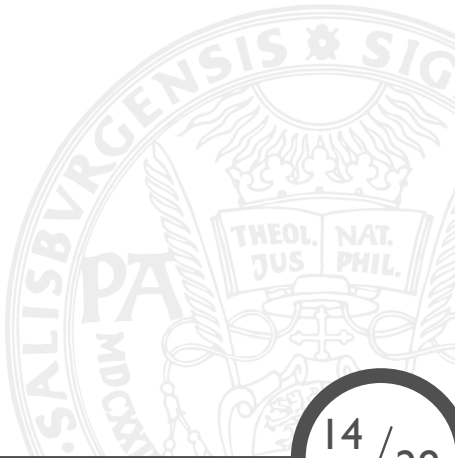
Light coding principles – What is the goal?

Goal of pattern design: decodable in the presence of non-idealities

- Assume that the projected pattern has $N_R \times N_C$ pixel p_A^i
- Active triangulation requires one codeword per pixel
- The more codewords are different $\rightarrow$ better robustness

In a calibrated + rectified setup ...

- conjugated points lie on horizontal lines
- $\rightarrow$ coding problem independent for each row
- we can use a small number of codewords



Structured Light Imaging

Light coding principles

In one row, there are $N := N_C^A$ pixel $p_A^1, \ldots, p_A^N$ to be encoded with codewords $w_1, \ldots, w_N$.

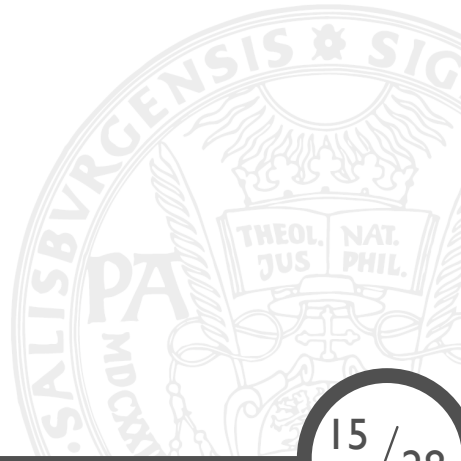
Structured Light Imaging

Light coding principles

In one row, there are $N := N_C^A$ pixel $p_A^1, \ldots, p_A^N$ to be encoded with codewords $w_1, \ldots, w_N$.

How is this implemented?

- Projector with n_P different illumination patterns



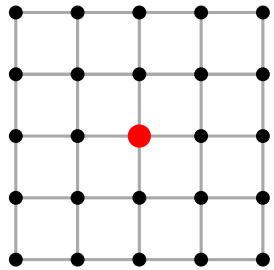
Structured Light Imaging

Light coding principles

In one row, there are $N := N_C^A$ pixel $p_A^1, \dots, p_A^N$ to be encoded with codewords $w_1, \dots, w_N$.

How is this implemented?

- Projector with n_P different illumination patterns
- E.g., in a window with n_W pixel



we have $n_P^{n_W}$ possible configurations

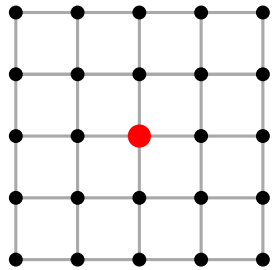
Structured Light Imaging

Light coding principles

In one row, there are $N := N_C^A$ pixel $p_A^1, \dots, p_A^N$ to be encoded with codewords $w_1, \dots, w_N$.

How is this implemented?

- Projector with n_P different illumination patterns
- E.g., in a window with n_W pixel



  we have $n_P^{n_W}$ possible configurations
- N of them need to be chosen to encode the pixel!

Structured Light Imaging

Light coding principles

Remember,

$$\underbrace{p_A = [u_A, v_A]^\top}_{\text{to be projected}} \rightarrow \underbrace{P = [x, y, z]^\top}_{\text{scene point}} \rightarrow \underbrace{p_C = [u_A + d, v_A]^\top}_{\text{camera point}}$$

where $\rightarrow$ denotes "projected to"

Structured Light Imaging

Light coding principles

Remember,

$$\underbrace{p_A = [u_A, v_A]^\top}_{\text{to be projected}} \rightarrow \underbrace{P = [x, y, z]^\top}_{\text{scene point}} \rightarrow \underbrace{p_C = [u_A + d, v_A]^\top}_{\text{camera point}}$$

where $\rightarrow$ denotes "projected to"

Projection introduces a horizontal shift d, inverse proportional to the estimated depth z, i.e.,

$$d = \frac{b|f|}{z}$$

Structured Light Imaging

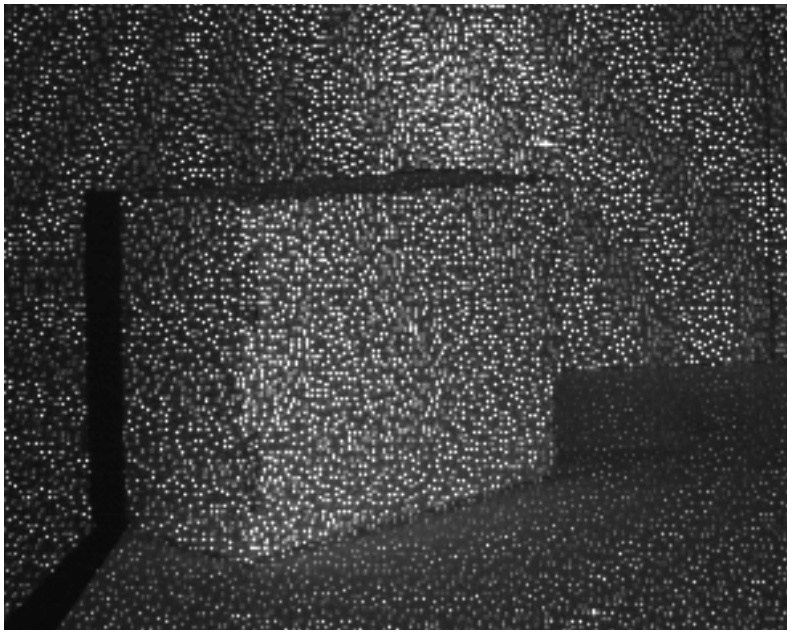
Light coding principles – Artificats

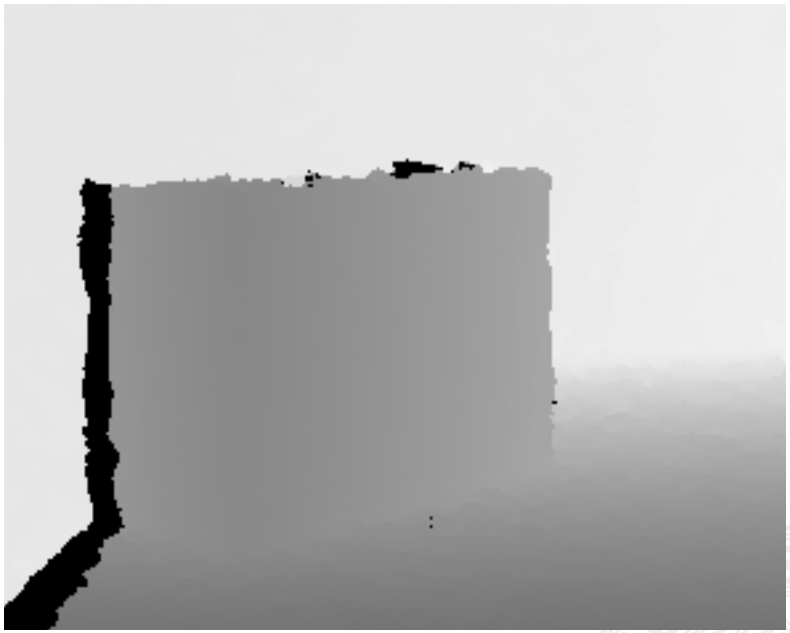
- Perspective distortion
- Color/Gray-level distortion
- Projector/Camera non-idealities
- Projector/Camera noise
- External illumination
- Occlusions

Structured Light Imaging

Light coding techniques – Artifacts

Projected pattern + depth map of slanted surface

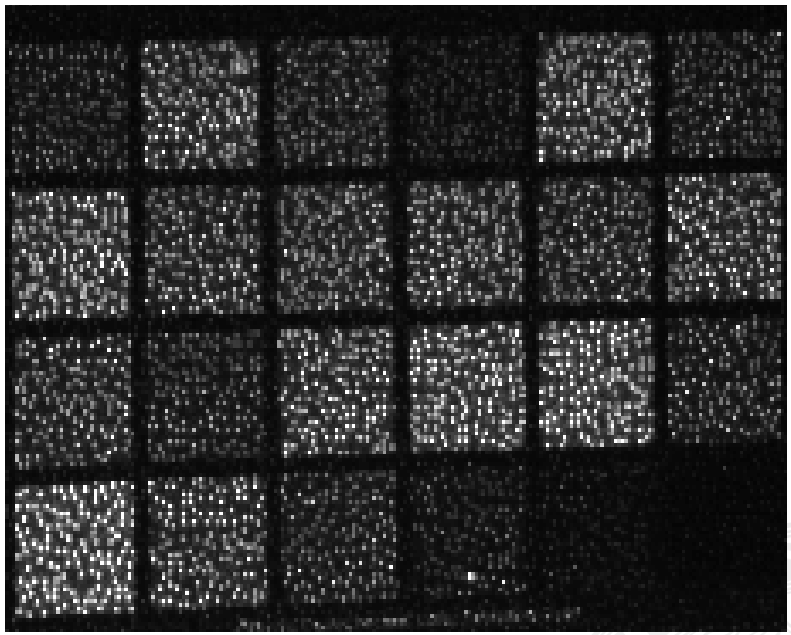




Structured Light Imaging

Light coding techniques – Artifacts

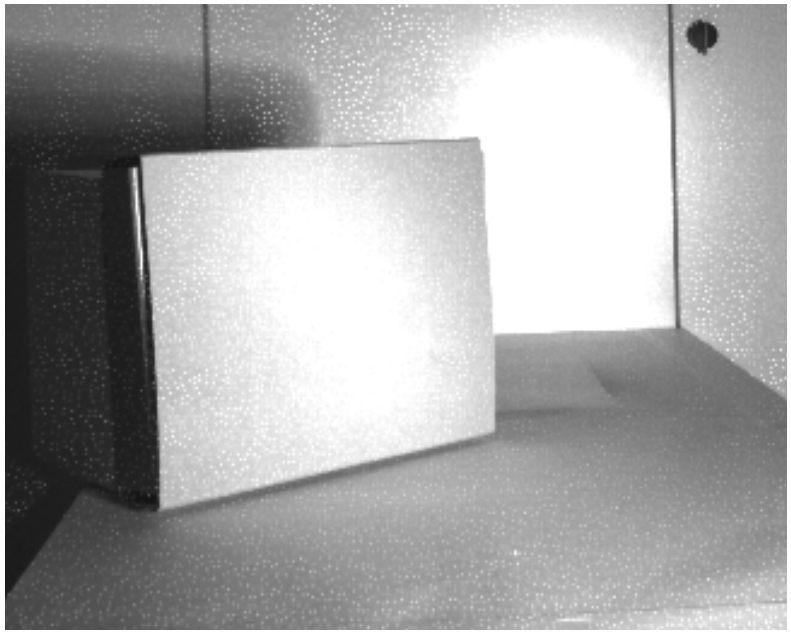
Projected pattern of color checkerboard (on left)



Structured Light Imaging

Light coding techniques – Artifacts

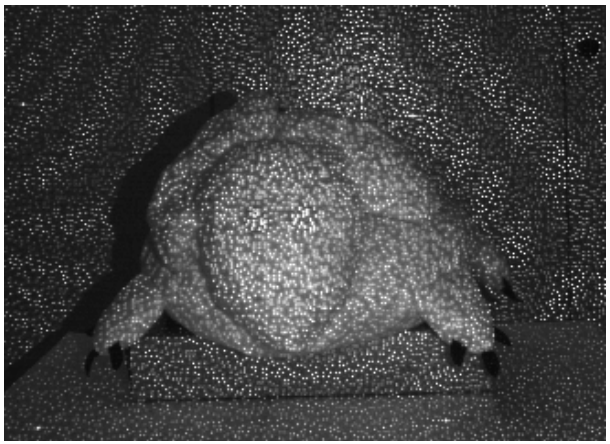
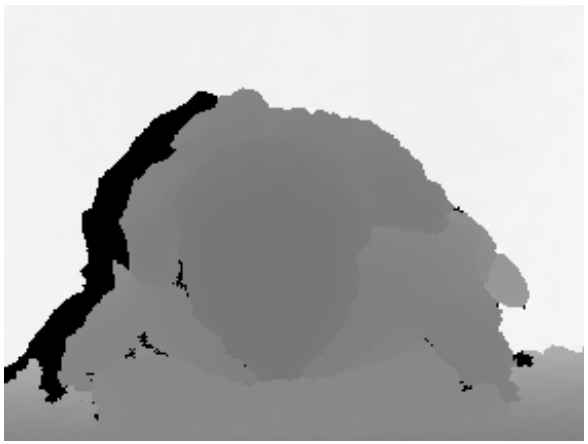
Effect on a strong external light source. The IR image saturates (relative to the strongest reflections) $\rightarrow$ no depth values (black regions in right image)



Structured Light Imaging

Light coding techniques – Artifacts

Occlusion (look at the left ear)

Structured Light Imaging

Light coding principles – Coding schemes

Two fundamental design considerations:

1. What codeword to assign to pixel p_A?

   The codeword corresponds to a pattern to be projected on a window centered at p_A.

Structured Light Imaging

Light coding principles – Coding schemes

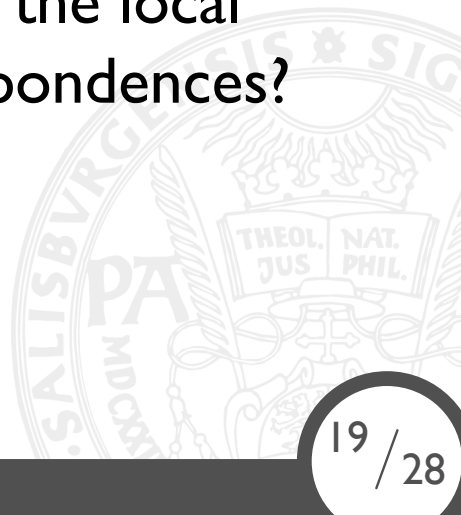
Two fundamental design considerations:

1. What codeword to assign to pixel p_A?

   The codeword corresponds to a pattern to be projected on a window centered at p_A.

2. What codeword to assign to pixel p_C?

   How do we identify the codeword most similar to the local pattern distribution around p_C to establish correspondences?

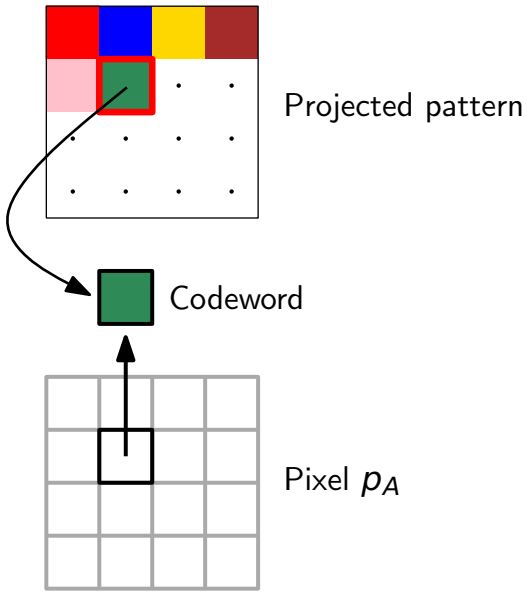


Structured Light Imaging

Light coding techniques – Coding schemes

What codeword should we assign to a pixel p_A?

1. Direct coding

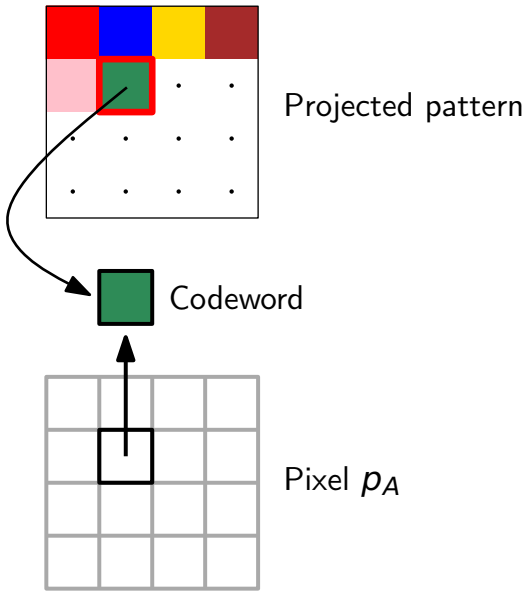


Structured Light Imaging

Light coding techniques – Coding schemes

What codeword should we assign to a pixel p_A?

1. Direct coding

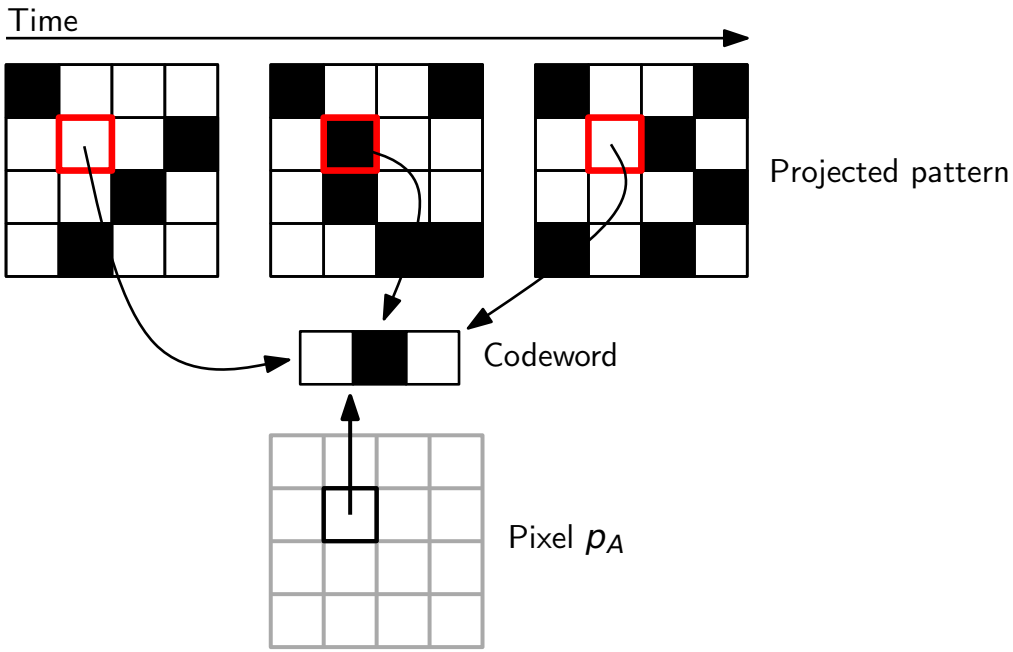


The codeword at p_A is the pattern value at that pixel $\rightarrow$ up to n_P codewords, since $n_W = 1$ (i.e., the "window" is just 1 pixel)

Structured Light Imaging

Light coding techniques – Coding schemes

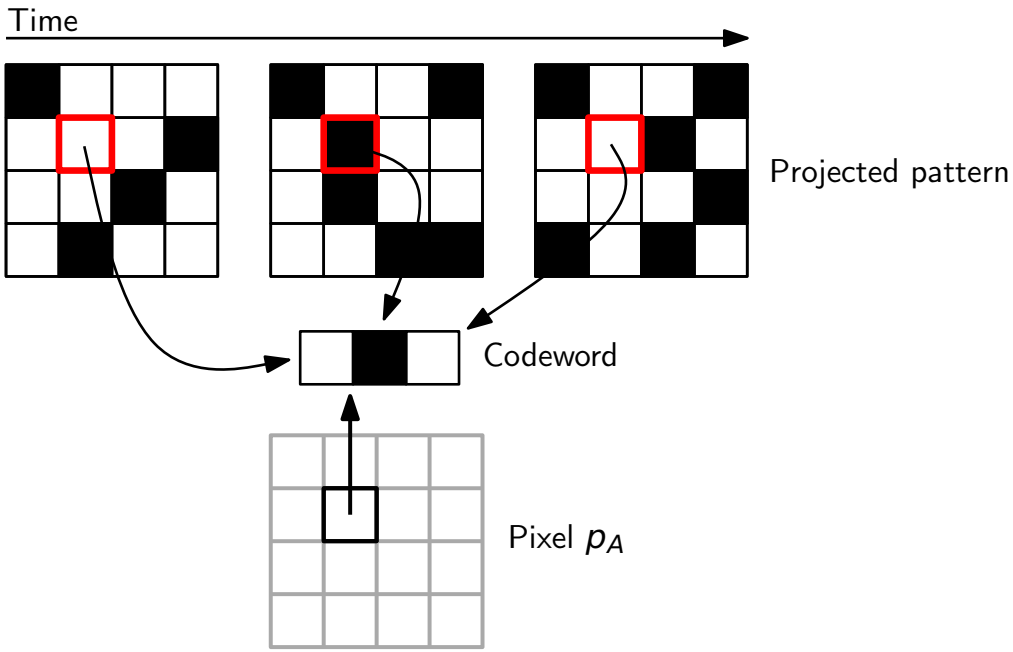
2. Time-multiplexed coding



Structured Light Imaging

Light coding techniques – Coding schemes

2. Time-multiplexed coding

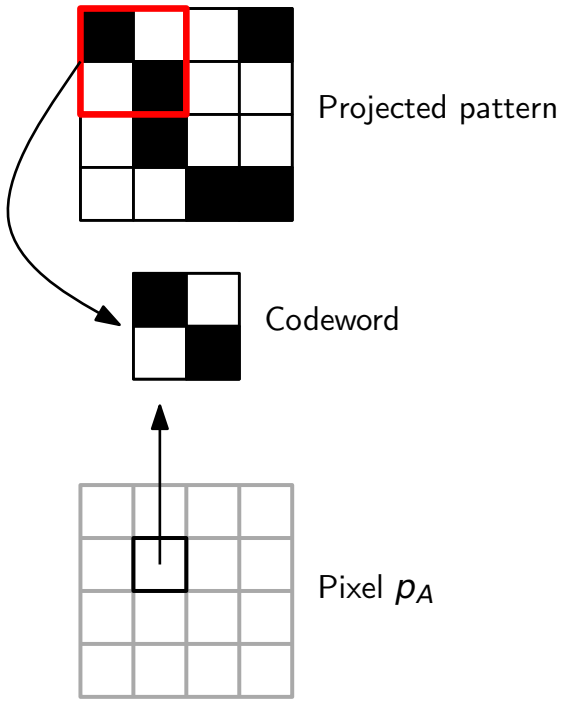


The codeword at pixel p_A is the sequence of projected patterns at that pixel $\rightarrow$ up to n_P^T codewords (pattern is projected T times).

Structured Light Imaging

Light coding techniques – Coding schemes

3. Spatial-multiplexed coding

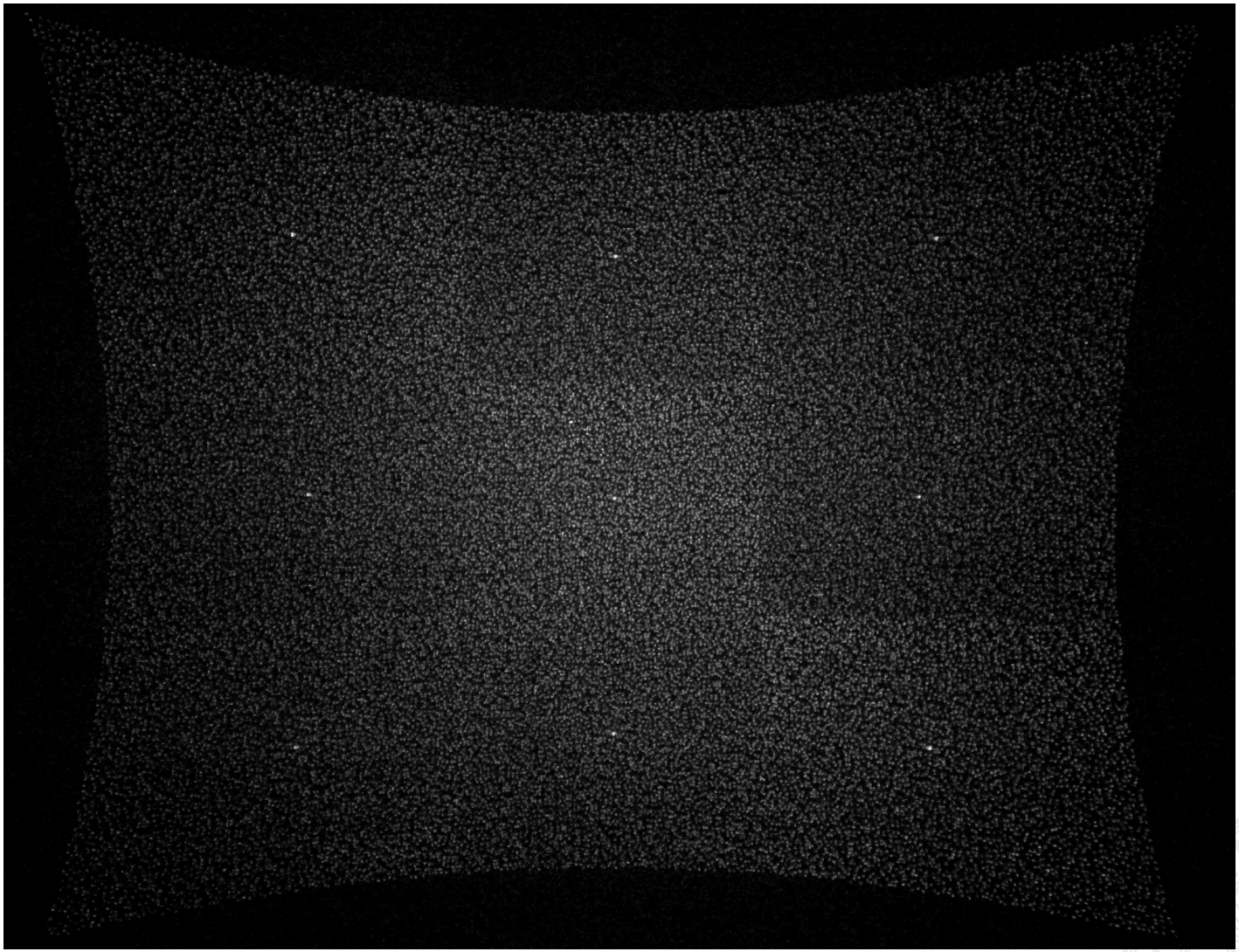


The codeword at pixel p_A is the spatial pattern distribution in a window with n_W pixel centered at $p_A \rightarrow$ up to $n_P^{n_W}$ codewords.

Structured Light Imaging

Light coding principles – Pixel matching on the example of Microsoft's Kinect™

- Spatial-multiplexed coding (ideal for dynamic scenes)
- A and C might have different spatial resolutions (undisclosed)
- The projected pattern is uncorrelated across each row
- Reverse engineering suggests a window of either 7×7 or 9×9



Pattern projected by the Kinect™ (on flat surface)

Structured Light Imaging

Light coding principles – Pixel matching on the example of Microsoft's Kinect™

What does "uncorrelated across each row" mean ?

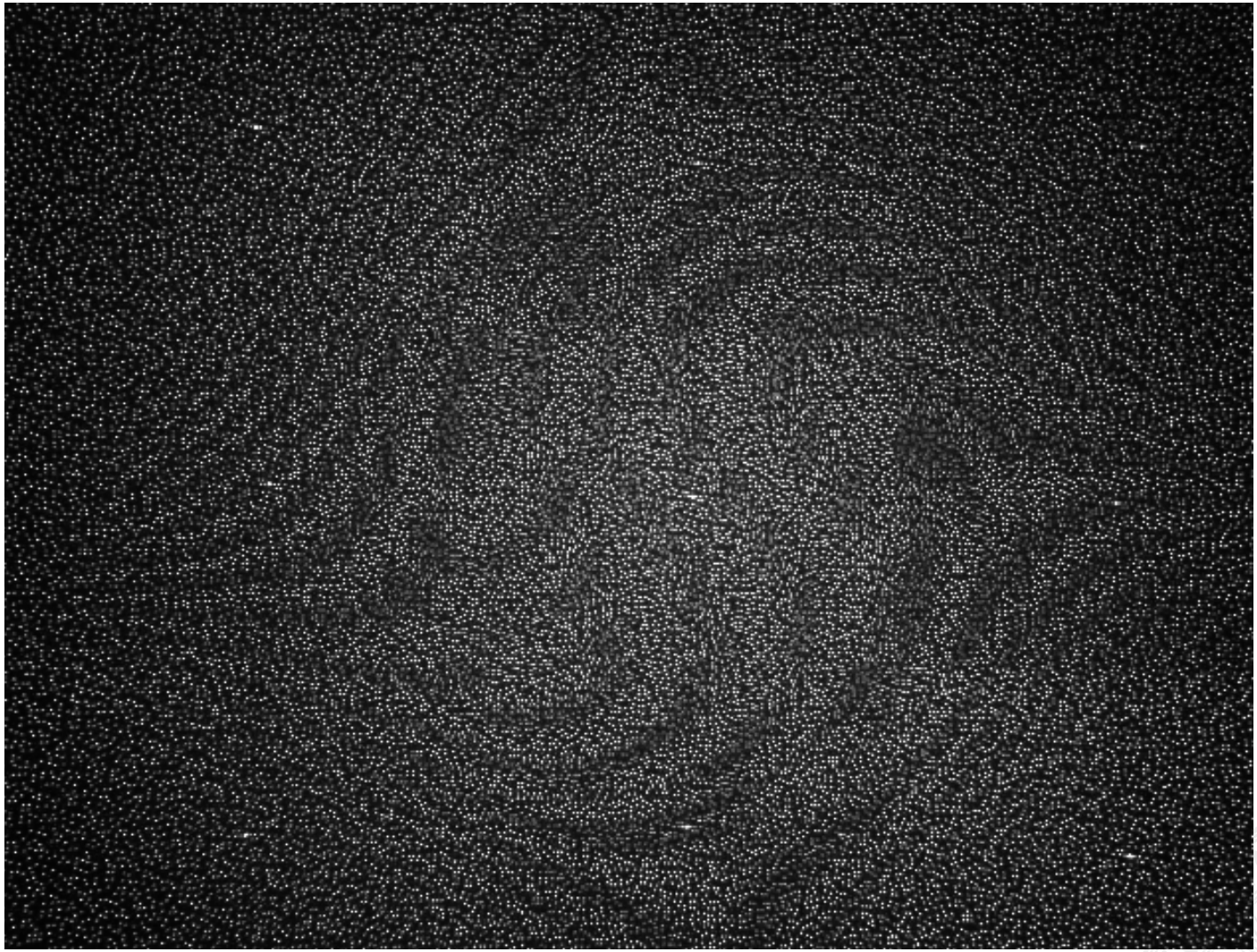
Let $s(u, v)$ be the projected pattern and $W(u_A, v_A)$ the support of the spatial multiplexing window, centered at p_A. Then

$$\underbrace{C(u^i, u^j, v_i)}_{\text{horizontal covariance between supports centered at } p_A^i \text{ and } p_A^j}$$

$$= \sum_{(u,v)\in W(u_A,v_A)} [s(u - u_A^i, v - v_A^i) - \overline{s}(u_A^i, v_A^i)][s(u - u_A^j, v - v_A^j) - \overline{s}(u_A^j, v_A^j)]$$

is 1 if $i = j$ and 0 otherwise (in the ideal case), where

$$\overline{s}(u_A, v_A) = \sum_{(u,v)\in W(u_A,v_A)} s(u - u_A, v - v_A) \ .$$



Pattern acquired by the Kinect™ IR camera

Structured Light Imaging

Light coding principles – Pixel matching on the example of Microsoft's Kinect™

Pixel matching: Ideally, for each pixel p_C^j of I_K and each pixel p_A^i in the same row, we obtain a unique covariance peak for an actual couple of conjugate points (there will be noise in real life :)

Slide credits / Literature

Most of the material presented in this lecture is either taken from the textbook of Dal Mutto et al.,[1] online at

http://freia.dei.unipd.it/nuovo/Papers/ToF-Kinect-book.pdf

and the PhD thesis of O. Elkhalili .[2]

[1] C. Dal Mutto, P. Zanuttigh, and G.M. Cortelazzo. *Time-of-Flight Cameras and Microsoft Kinect – A user perspective on technology and applications*. Springer, 2013.

[2] Omar Elkhalili. "Entwicklung von optischen 3D CMOS-Bildsensoren auf der Basis der Pulslaufzeitmessung". PhD thesis. Universität Duisburg-Essen, 2005.