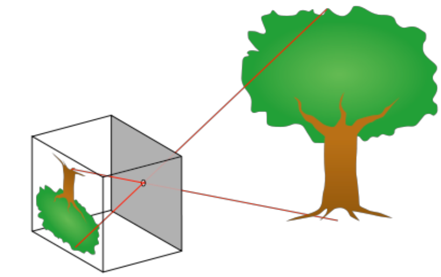
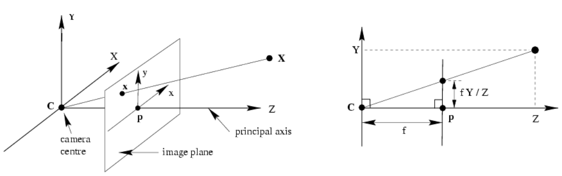
Image formation: pinhole model



(a) Texture map

Pinhole camera



space point image point

$$(X,Y,Z)^T \mapsto \mathbf{x} = (x,y)^T = (fX/Z, fY/Z)^T$$

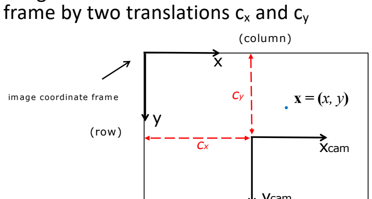
both in camera coordinate frame!

Terms and definitions

- **Camera centre**: point (pinhole) through which light enters the camera
- **Image plane**: (virtual) plane in front of the camera centre where image is formed/sensed
- **Focal length (f)**: distance between camera centre and the image plane
- **Optical/principal axis (z)**: line through camera centre and perpendicular to the image plane
- **Camera coordinate frame**
  - origin: camera centre
  - z-axis: along optical axis
  - x-y plane: parallel to image plane with the horizontal axis being x-axis
- **Space point**: a point in space in the camera coordinate frame
- **Image point**: the projection of a space point on the image plane
- **Image coordinate frame:** shares x and y-axis directions with camera coordinate frame but whose origin is the upper left corner of the image plane.

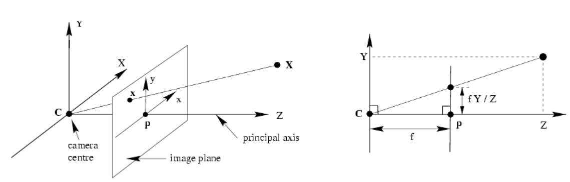
Image vs camera coordinate frame

- Image coordinate frame is related to the camera coordinate frame by two translations c_x and c_y



- In addition, image point $\mathbf{x}$ is mm in the camera coordinate frame but in pixels in the image coordinate frame.

Pinhole camera model



Therefore, image point in image coordinate frame is given by:

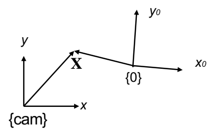
$$(f_x X/Z + c_x,\ f_y Y/Z + c_y)$$

(c_x, c_y) are the image centre, and (f_x, f_y) are focal lengths, which are called **intrinsic parameters** of the camera.

Extrinsics

- If the world coordinate frame {0} is defined with respect to the camera coordinate frame ({cam}), by R and t,

$$\mathbf{X}_c = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \mathbf{X}_0$$



- $[R,\ t]$ are called the **extrinsic parameters** of the camera and K the intrinsic parameters.

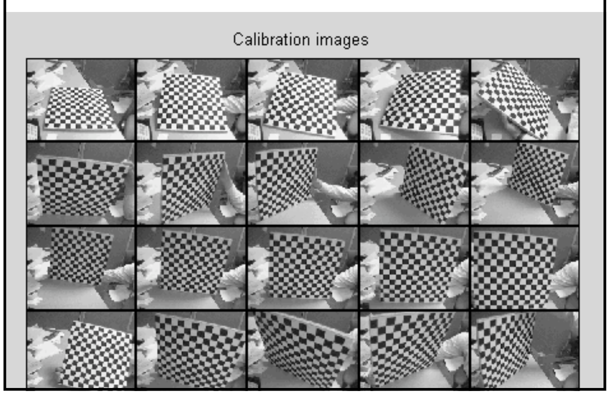
12

Summary of camera parameters

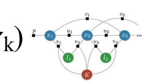
- Intrinsics: K
  - Focal lengths (2)
  - Image center (2)
  - Skew coefficient (1) (angle between x and y axes)
  - Distortion coefficients (3) (radial distortion)
- Extrinsics: $R,\ t$
  - Rotation (3), R
  - Translation (3), t
- Intrinsics are calibrated once whereas extrinsics may change whenever the camera (robot) moves.

13

Camera Calibration: Identify Intrinsics



Measurement function $h_k(\chi_k)$

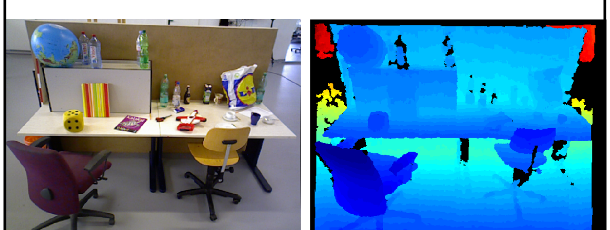


- χ_k: camera/robot pose x_i (2D or 3D) and landmark position $l_j = (X,\ Y,\ Z)$ (3D)
- z_k: image point in image coordinate frame $(x,\ y)$, e.g., $(56,\ 73)$ pixels
- $h_k(\chi_k)$ computes the expected image point in:

$$\mathrm{p}(z_k|\mathcal{X}_k) \propto \exp\left(-\frac{1}{2}||h_k(\mathcal{X}_k) - z_k||^2_{\Omega_k}\right) \qquad (3)$$

based on: $(f_x X/Z + c_x,\ f_y Y/Z + c_y)$, with appropriate coordinate transformation

RGB-D: Xtion Pro Live/Kinect/Astra



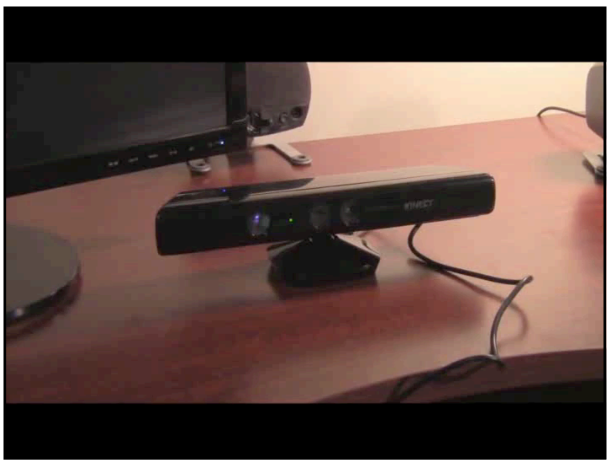
(a) Texture map (b) Depth map

Xtion Pro Live

- Introduction
  - How it works (video)
  - Structured light (video)
- Hardware
  - RGB
  - Depth sensing by structured light

Introduction

- Patented technology by an Israeli Company (PrimeSense)
- Developed for gaming
- Microsoft, Asus and Orbbecc licensed technology



Specification

PRODUCT SPECIFICATION

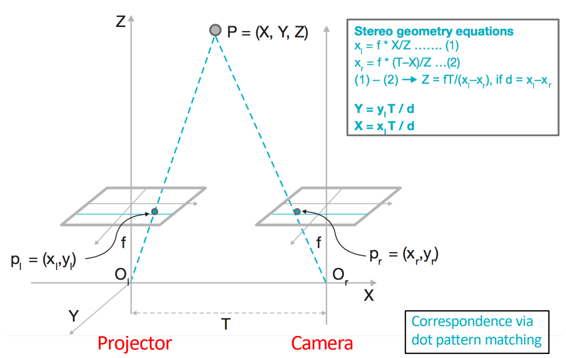
Property	PrimeSensor Spec
Field of View (Horizontal, Vertical, Diagonal)	58° H, 40° V, 70° D
Depth image size	VGA (640x480)
Spatial x/y resolution (@2m distance from sensor)	3mm
Depth z resolution (@2m distance from sensor)	1cm
Maximal image throughput (frame rate)	60fps
Average image latency in full VGA resolution	40msec
Operation range	0.8m - 3.5m

Property	PrimeSensor Spec
Color image size	UXGA (1600x1200)
Audio: built-in microphones	2 mics
Audio: digital inputs	4 inputs
Data interface	USB 2.0
Power supply	USB 2.0
Power consumption	2.25W
Dimensions (Width x Height x Depth)	14cm x 3.5cm x 5cm
Operation environment (every lighting condition)	indoor
Operating temperature	0°C - 40°C

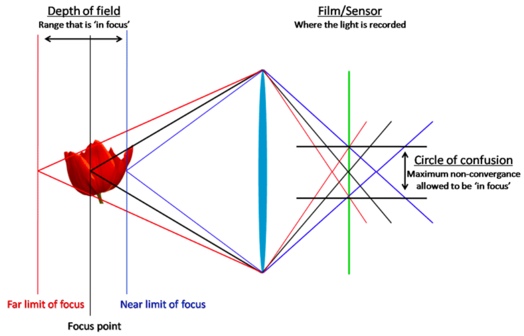
How it works

- The Kinect/Xtion Pro Live uses structured (IR) light principle to compute depth of a scene
- The depth computation is done by the PrimeSense Hardware built into the camera, and details are not available
- The basic principle, however, is well known: the depth from stereo triangular (and focus).

Depth from stereo



Depth from focus



Limitations of Xtion Pro Live

- Indoor only since it uses IR as projected light
- Interfere between multiple sensors in the same environment
- Limited FOV (58, 40, and 70 degrees respectively)
- Limited range (depth of field): 0.8 – 3.5 m
- Limited spatial resolution: 3 mm in x-y, 10 mm in z
- Communication bandwidth
- Power consumption

LiDAR vs. depth from RGB-D

- Lidar produces a point cloud: list of 3D points
- D channel of an RGB-D camera produces a depth map
- Given one, you can compute the other, i.e., depth image from point cloud and point cloud from depth image.
- The measurement function of a RGB-D camera is no different from that of a LiDAR, in terms of its depth value.