Time-of-Flight and Kinect Imaging

Victor Castaneda, Nassir Navab

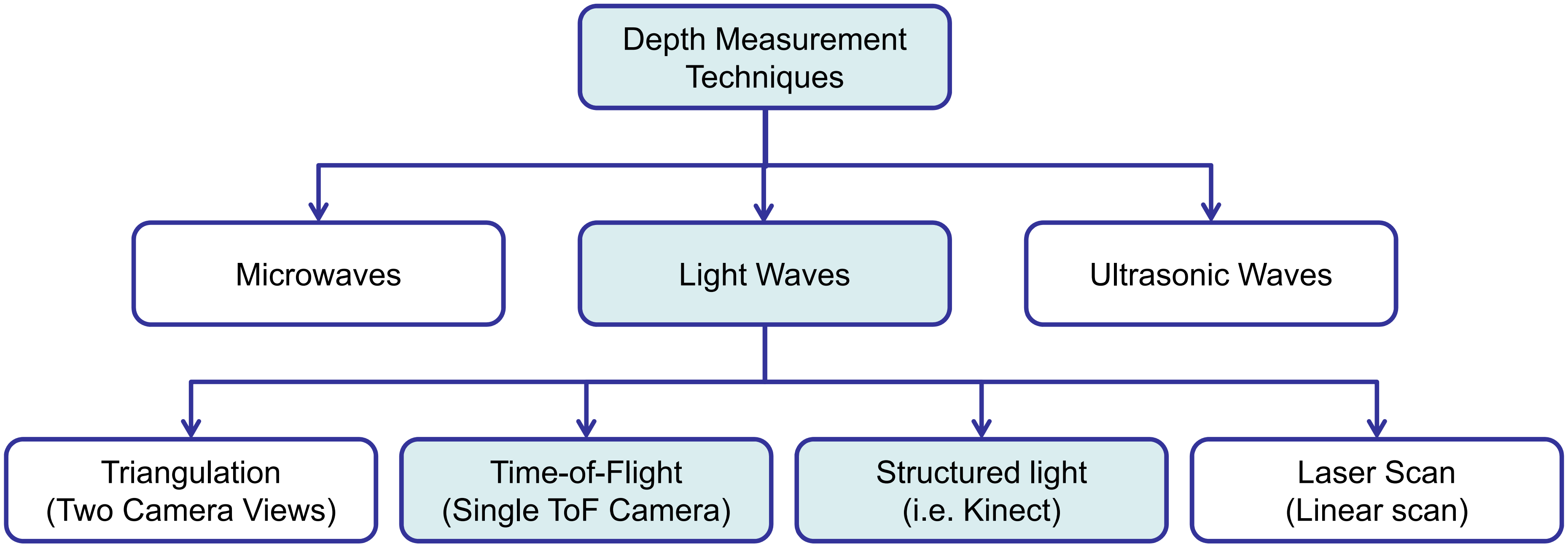
Kinect Programming for Computer Vision
Summer Term 2011 – 1.6.2011

Lecture Outline

1. Introduction and Motivation
2. Principles of ToF Imaging
3. Computer Vision with ToF Cameras
4. Principles of Kinect (Primesensor)
5. Case Studies

Introduction and Motivation

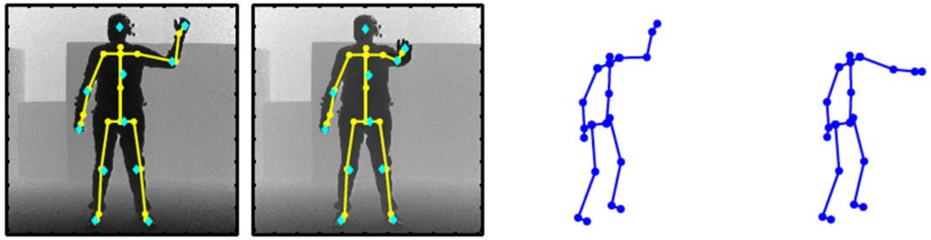
Classification of Depth Measurement Techniques



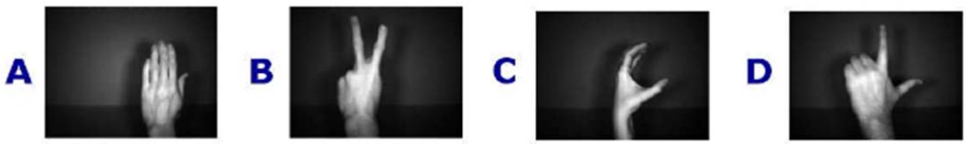
Introduction and Motivation

Applications for 3D Sensing

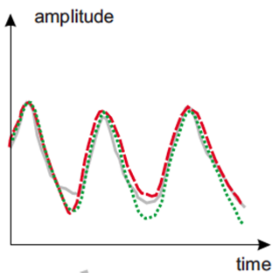
- Computer Vision
  - People and object tracking
  - 3D Scene reconstruction



- Interaction
  - Gesture-based user interfaces
  - Gaming/character animation



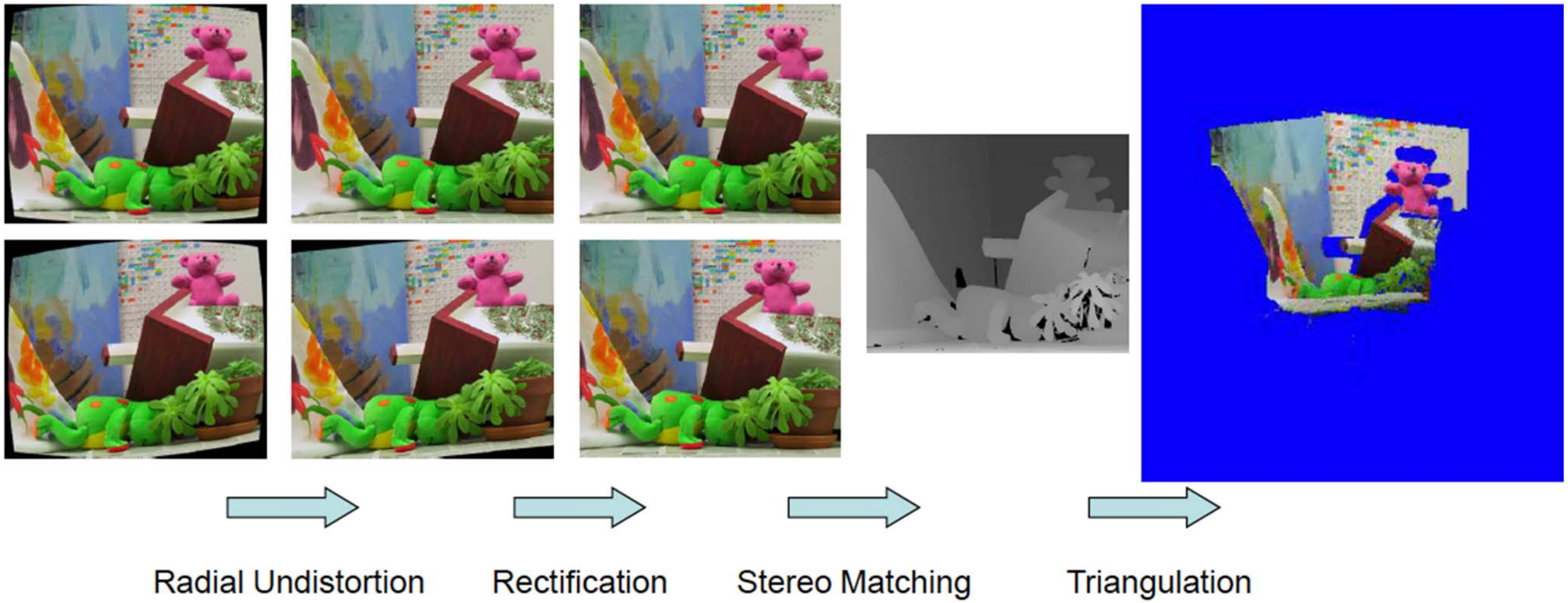
- Medical
  - Respiratory gating
  - Ambulatory motion analysis



Images from [8,12,14]

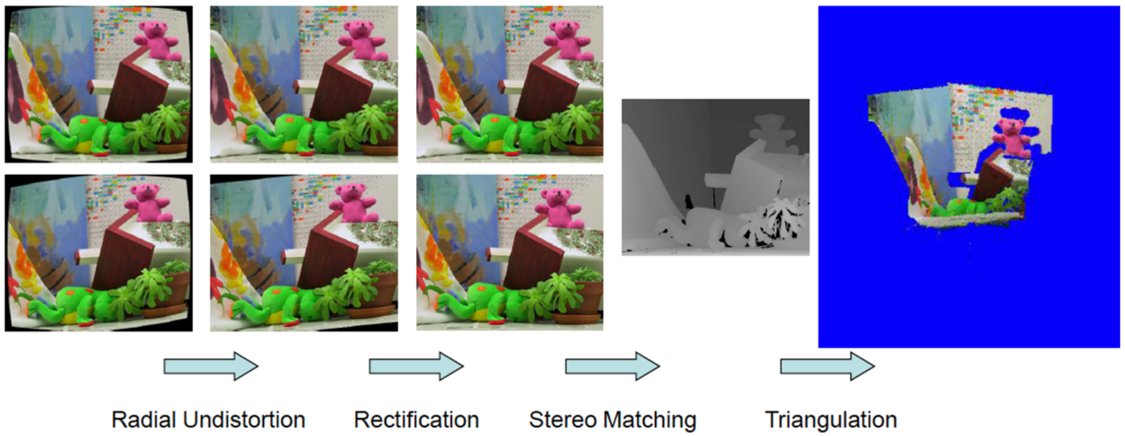
Introduction and Motivation

Depth Measurement Using Multiple Camera Views



Introduction and Motivation

Depth Measurement Using Multiple Camera Views

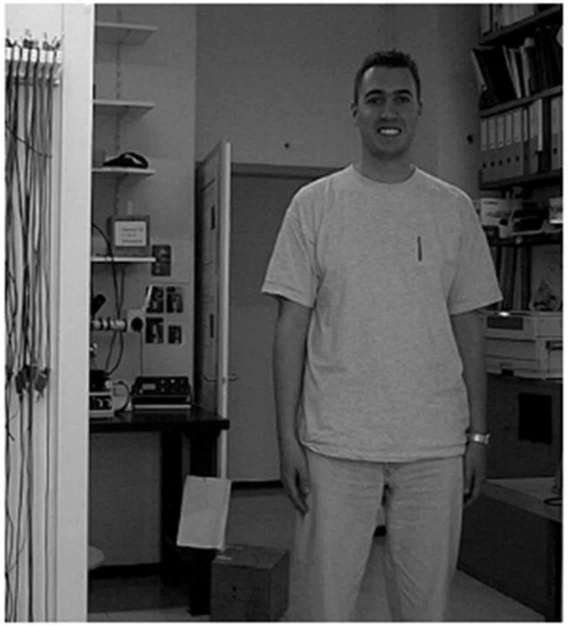


Disadvantages:

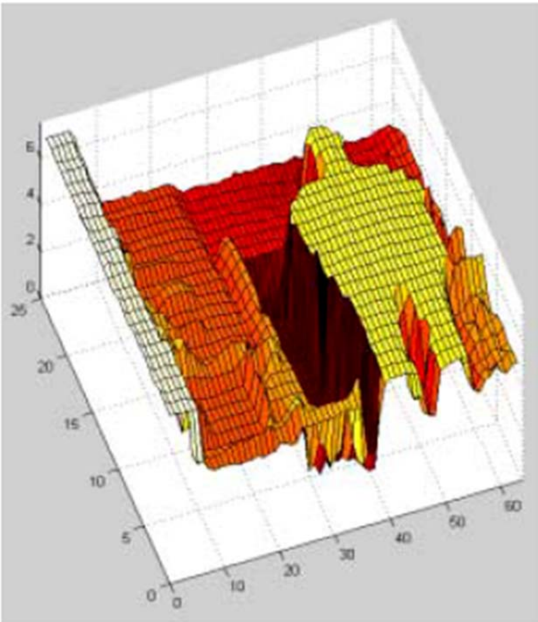
- At least two calibrated cameras required
- Multiple computationally expensive steps
- Dependence on scene illumination
- Dependence on surface texturing

Introduction and Motivation

Time-of-Flight (ToF) Imaging refers to the process of measuring the depth of a scene by quantifying the changes that an emitted light signal encounters when it bounces back from objects in a scene.



Regular Camera Image

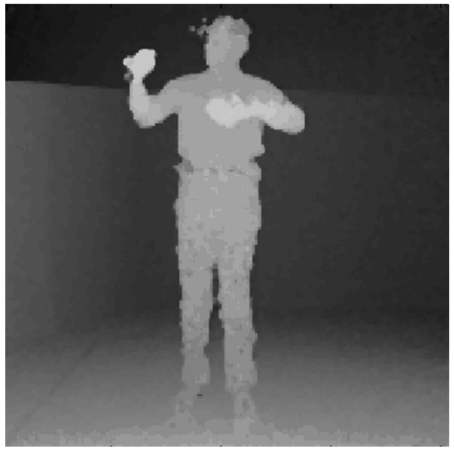


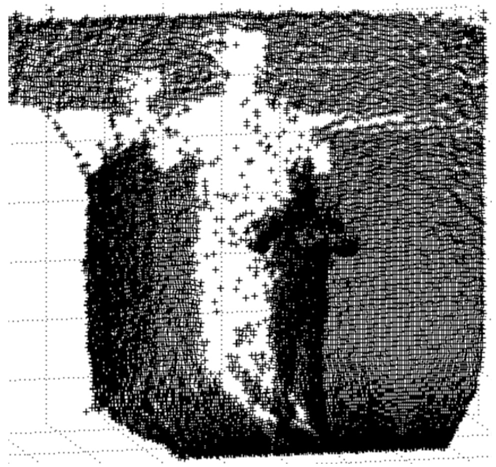
ToF Camera Depth Image

Images from [2]

Introduction and Motivation

Depth Measurement Using a ToF Camera

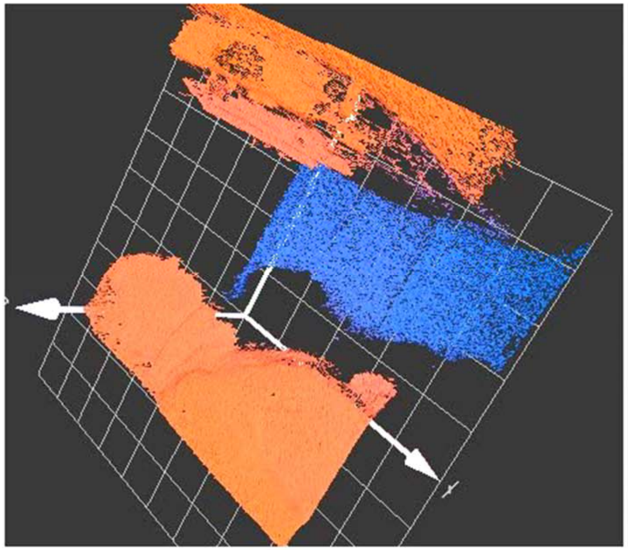




Advantages:

- Only one (specific) camera required
- No manual depth computation required
- Acquisition of 3D scene geometry in real-time
- Reduced dependence on scene illumination
- Almost no dependence on surface texturing

Introduction and Motivation



3D Reconstruction

ToF Amplitude Image

ToF Depth Image

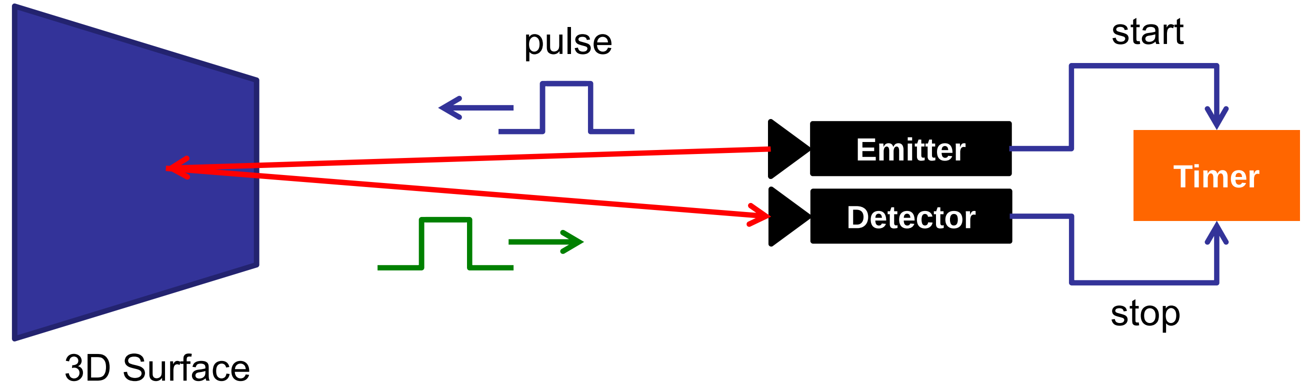
Lecture Outline

1. Introduction and Motivation
2. Principles of ToF Imaging
3. Computer Vision with ToF Cameras
4. Principles of Kinect (Primesensor)
5. Case Studies

Principles of ToF Imaging

Pulsed Modulation

- Measure distance to a 3D object by measuring the absolute time a light pulse needs to travel from a source into the 3D scene and back, after reflection
- Speed of light is constant and known, $c = 3 \cdot 10^8$m/s



Principles of ToF Imaging

Pulsed Modulation

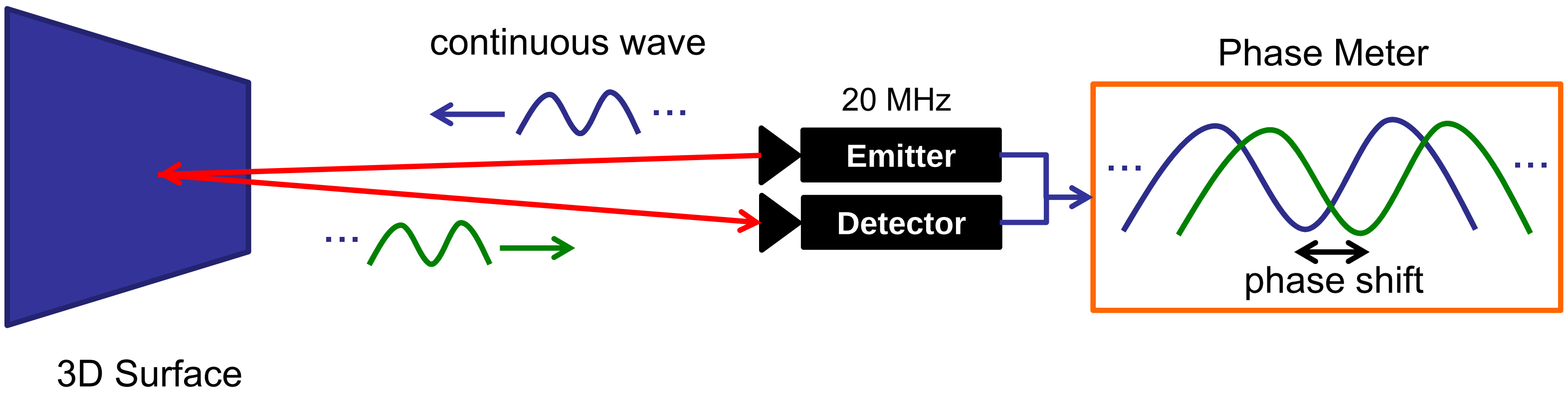
+ Advantages:
  - Direct measurement of time-of-flight
  - High-energy light pulses limit influence of background illumunation
  - Illumination and observation directions are collinear

- Disadvantages:
  - High-accuracy time measurement required
  - Measurement of light pulse return is inexact, due to light scattering
  - Difficulty to generate short light pulses with fast rise and fall times
  - Usable light sources (e.g. lasers) suffer low repetition rates for pulses

Principles of ToF Imaging

Continuous Wave Modulation

- Continuous light waves instead of short light pulses
- Modulation in terms of frequency of sinusoidal waves
- Detected wave after reflection has shifted phase
- Phase shift proportional to distance from reflecting surface



Principles of ToF Imaging

Continuous Wave Modulation

- Retrieve phase shift by demodulation of received signal
- Demodulation by cross-correlation of received signal with emitted signal
- Emitted sinusoidal signal:

$$g(t) = \cos(\omega t)$$

ω: modulation frequency

- Received signal after reflection from 3D surface:

$$s(t) = b + a\cos(\omega t + \phi)$$

b: constant bias
a: amplitude
ϕ: **phase shift**

- Cross-correlation of both signals:

$$c(\tau) = s * g = \int_{-\infty}^{\infty} s(t) \cdot g(t+\tau)dt$$

τ: offset

Principles of ToF Imaging

Continuous Wave Modulation

- Cross-correlation function simplifies to

$$c(\tau) = \frac{a}{2}\cos(\omega\tau + \phi) + b$$

b: constant bias
a: amplitude
ϕ: **phase shift**
τ: internal offset

- Sample $c(\tau)$ at four sequential instants with different phase offset τ:

$$A_i = c(i \cdot \pi/2), \quad i = 0, \ldots, 3$$

- Directly obtain sought parameters:

$$\phi = \text{arctan2}(A_3 - A_1, A_0 - A_2)$$

$$a = 1/2\sqrt{(A_3 - A_1)^2 + (A_0 - A_2)^2}$$

distance:

$$\boxed{\Rightarrow d = \frac{c}{4\pi\omega}\phi}$$

Principles of ToF Imaging

Continuous Wave Modulation

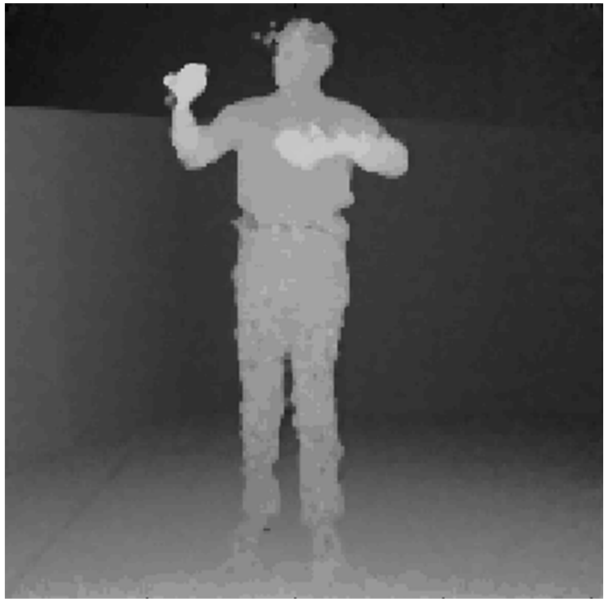
+ Advantages:
- Variety of light sources available as no short/strong pulses required
- Applicable to different modulation techniques (other than frequency)
- Simultaneous range and amplitude images

- Disadvantages:
- In practice, integration over time required to reduce noise
- Frame rates limited by integration time
- Motion blur caused by long integration time

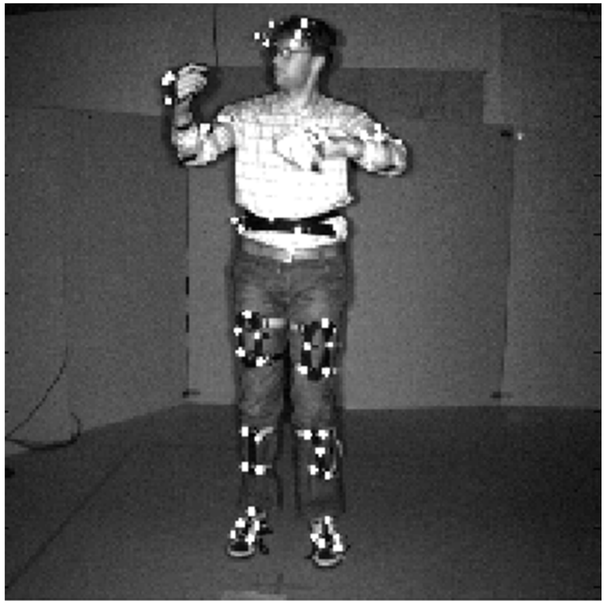
Principles of ToF Imaging

Continuous Wave Modulation

- Simultaneous availability of (co-registered) range and amplitude images



Depth Image



Amplitude Image

Principles of ToF Imaging

Example Device: PMDVision CamCube

- Near-infrared light (700-1400 nm)
- Continous wave modulation
- Sinusoidal signal

- Resolution: 204x204 pixels
- Standard lens, standard calibration
- Frame rate: 20 fps

- Multiple camera operation by using different modulation frequencies

Image from [3]

Lecture Outline

1. Introduction and Motivation
2. Principles of ToF Imaging
3. Computer Vision with ToF Cameras
4. Principles of Kinect (Primesensor)
5. Case Studies

Computer Vision with ToF Cameras

Measurement Errors and Noise

Systematic distance error

- Perfect sinusoidal signals hard to achive in practice
- Depth reconstructed from imperfect signals is erroneous
- Solution 1: camera-specific calibration to know distance error
- Solution 2: alternative demodulation techniques not assuming perfect sinusoidal signals

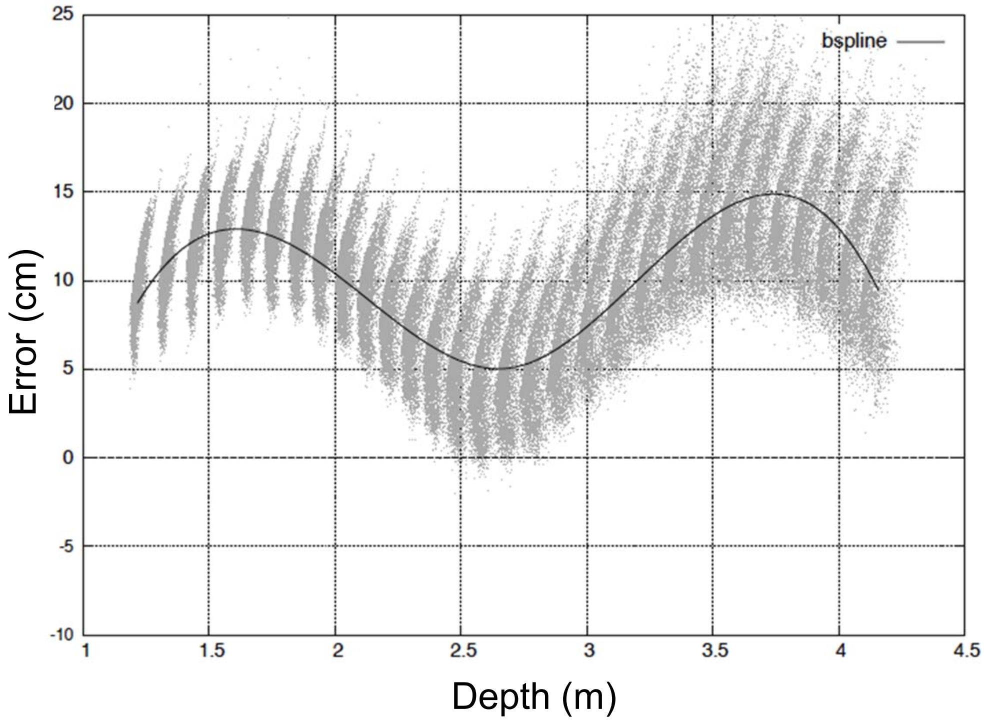


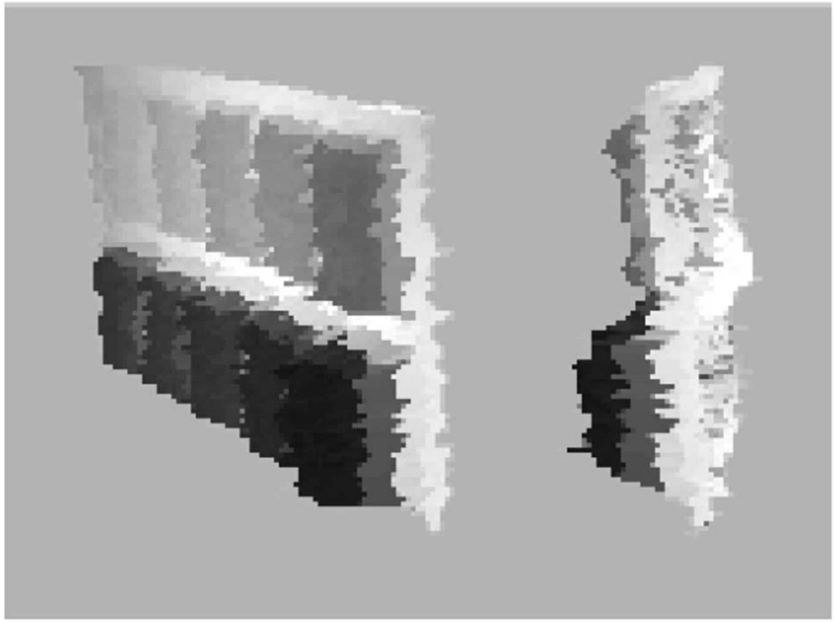
Image from [1]

Computer Vision with ToF Cameras

Measurement Errors and Noise

Intensity-related distance error

- Computed distance depending on amount of incident light
- Inconsistencies at surfaces with low infrared-light reflectivity
- Correction by means of corresponding amplitude image



Depth images of planar object with patches of different reflectivity

Image from [1]

Computer Vision with ToF Cameras

Measurement Errors and Noise

Depth inhomogeneity

- Current ToF cameras have low pixel resolution
- Individual pixels get different depth measurements
- Inhomogeneous
- „Flying pixels“, especially at object bondaries
- Correction: discard pixels along rays parallel to viewing direction

Red circles: „flying pixels“

Computer Vision with ToF Cameras

Measurement Errors and Noise

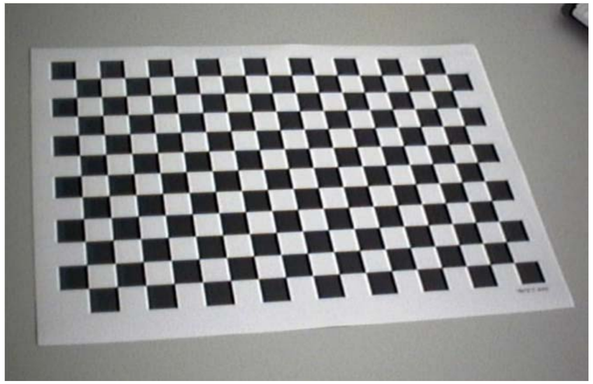
Light interference effects

- Signal received on detector can be mixed with signals that were reflected in the scene multiple times (instead of direct reflection)
- Emitted light waves can be attenuated and scattered in the scene
- Interference by other sources of near-infrared light (e.g. sunlight, infrared marker-based tracking systems, other ToF cameras)

Computer Vision with ToF Cameras

Geometric Calibration of ToF Cameras

- Standard optics used in commercial ToF cameras
- Use ToF amplitude image for calibration
- Standard calibration procedure for camera intrinsics
  - $f_x = fm_x$: focal length in terms of pixel dimensions (x)
  - $f_y = fm_y$: focal length in terms of pixel dimensions (y)
  - c_x: principal point (x)
  - c_y: principal point (y)
  - Lens distortion parameters
- Typical approach:
  - checkerboard calibration pattern
  - World-to-image point correspondences
  - Linear estimation of intrinsic/extrinsic parameters
  - Non-linear optimization



Computer Vision with ToF Cameras

Extraction of Metric 3D Geometry from ToF Data

- ToF data: depth *d* in meters for every pixel location $\mathbf{x} = (x, y)^\top$
- Desired data: 3D coordinates $\mathbf{X} = (X, Y, Z)^\top$ for every pixel

- Write image coordinates in homogeneous notation (x,y,1)
- Apply inverse of intrinsic parameters matrix K to points

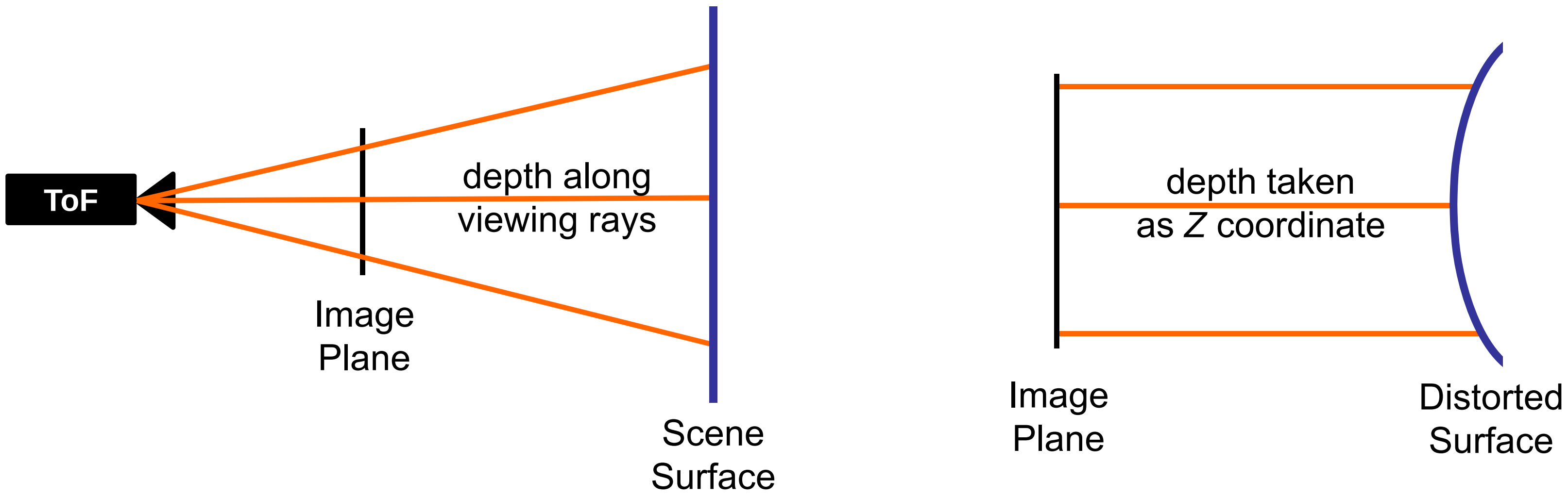
$$\mathbf{x} = \mathbf{P}\mathbf{X} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{X}$$

$$\underbrace{\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} f_x X + c_x Z \\ f_y Y + c_y Z \\ Z \end{pmatrix}}_{\text{Camera projection 3D to 2D}} \qquad \underbrace{\begin{aligned} X &= \frac{(x - c_x)Z}{f_x} \\ Y &= \frac{(y - c_y)Z}{f_y} \end{aligned}}_{\text{Inverse relation for X and Y}}$$

Computer Vision with ToF Cameras

Extraction of Metric 3D Geometry from ToF Data

- Simply taking measured depth *d* as *Z* coordinate is not sufficient
- Depth is measured along rays from camera center through image plane



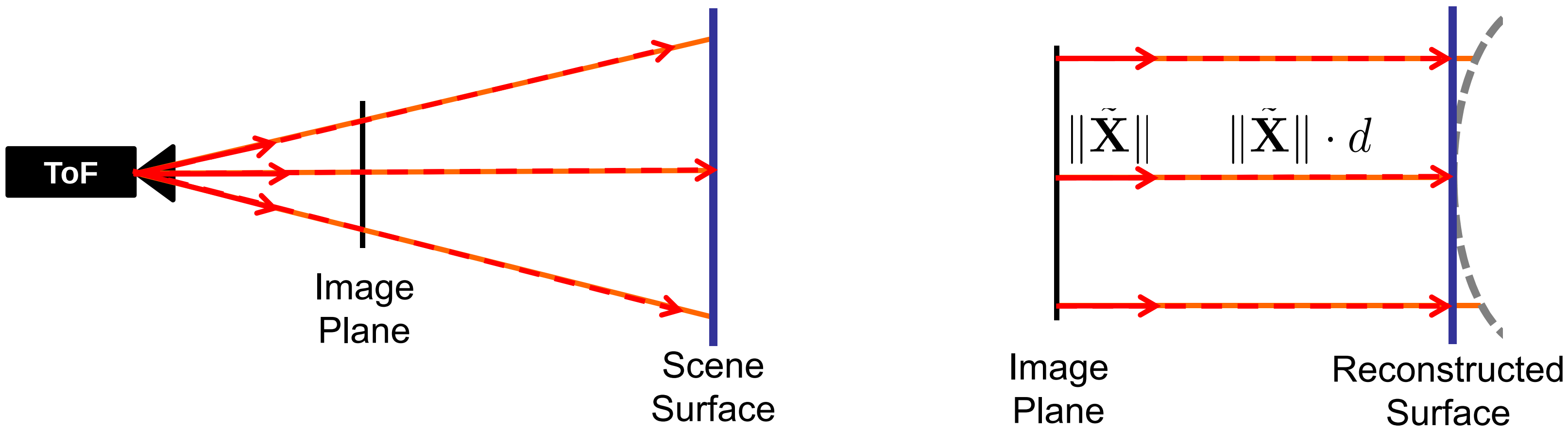
Computer Vision with ToF Cameras

Extraction of Metric 3D Geometry from ToF Data

- Ray from camera center into 3D scene:

$$\begin{pmatrix} (x-c_x)Z/f_x \\ (y-c_y)Z/f_y \\ Z \end{pmatrix}^\top \rightarrow \begin{pmatrix} (x-c_x)/f_x \\ (y-c_y)/f_y \\ 1 \end{pmatrix}^\top =: \tilde{\mathbf{X}}$$

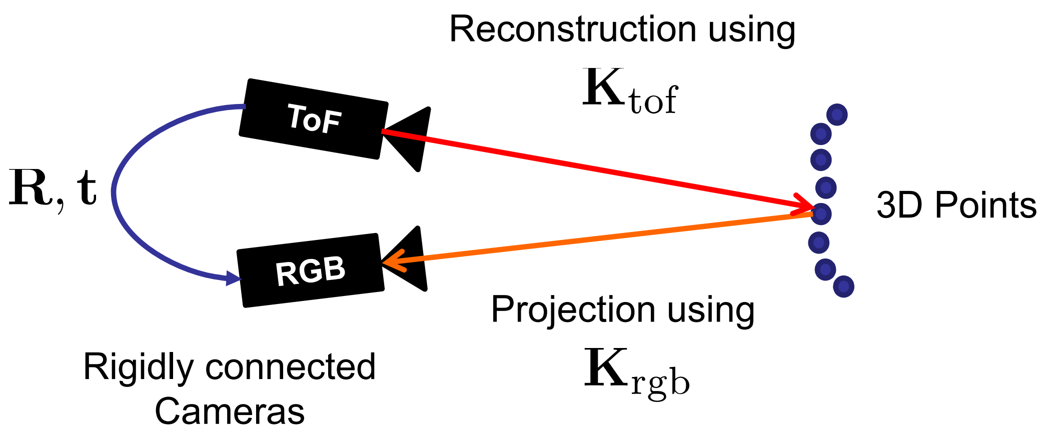
- Normalize to unit length (keep only direction), multiply with depth: $\mathbf{X} = \|\tilde{\mathbf{X}}\| \cdot d$



Computer Vision with ToF Cameras

Combining ToF with Other Cameras

- Additional, complementary information (e.g. color)
- Higher-resolution information (e.g. for superresolution)
- Example: combination with a high-resolution RGB camera
- Approach: Stereo calibration techniques, giving $\mathbf{R}, \mathbf{t}$ and $\mathbf{K}_{\text{tof}}, \mathbf{K}_{\text{rgb}}$



Lecture Outline

1. Introduction and Motivation
2. Principles of ToF Imaging
3. Computer Vision with ToF Cameras
4. Principles of Kinect (Primesensor)
5. Case Studies

Principles of Kinect (Primesensor)

Structured Light Imaging

- Project a known light pattern into the 3D scene, viewed by camera(s)
- Distortion of light pattern allows computing the 3D structure

Picture from Wikipedia

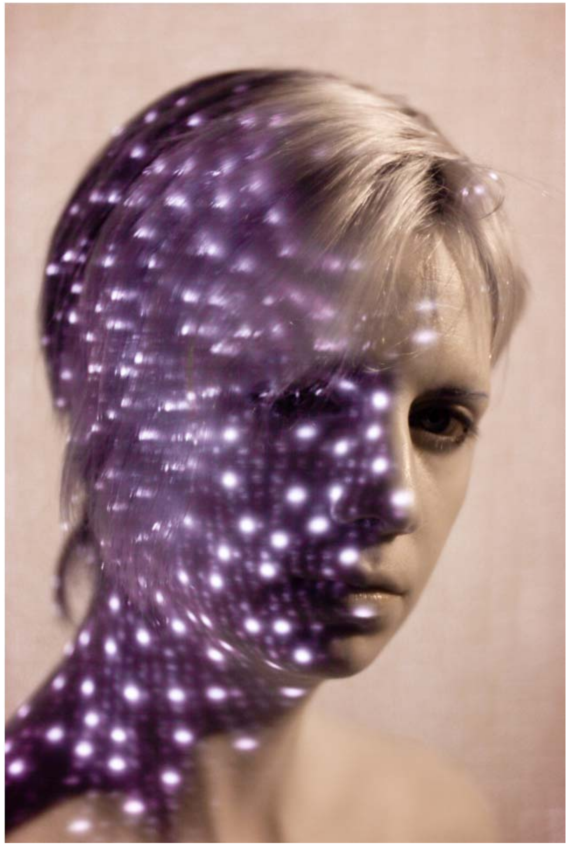
Principles of Kinect (Primesensor)

Structured Light Imaging types

- Time Multiplexing
- Direct coding
- **Spatial Neighborhood**

This coding has to be unique per position in order to recognize each point in the pattern.

Kinect uses pseudo random pattern.

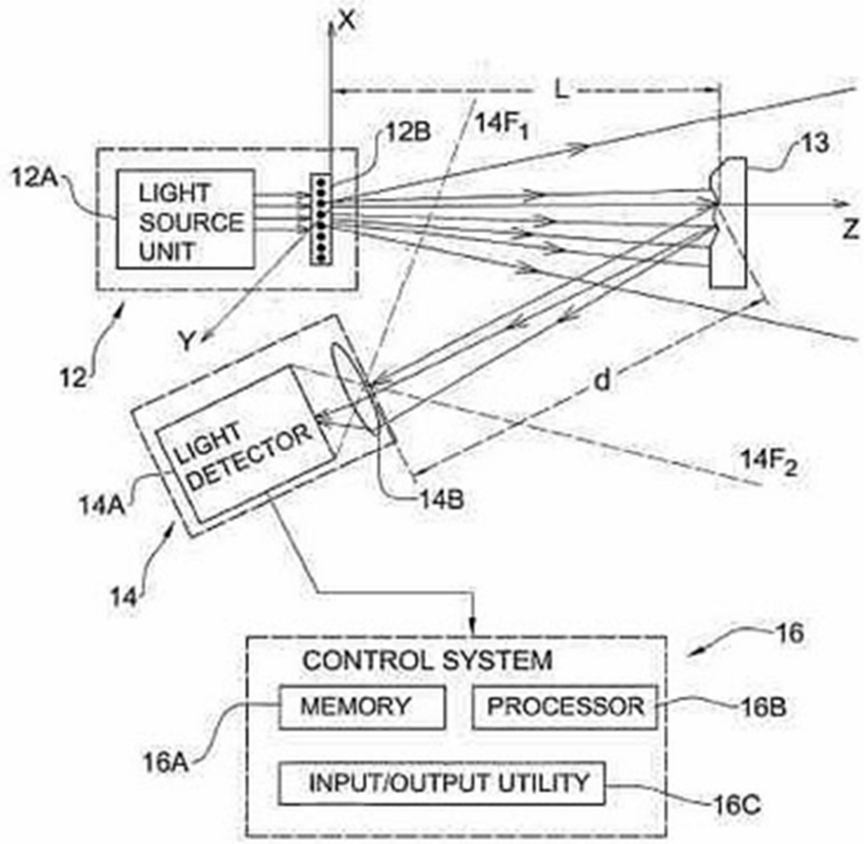


Picture from the Artist Audrey Penven

Principles of Kinect (Primesensor)

How Kinect works?

- Projects a known pattern (Speckles) in Near-Infrared light.
- CMOS IR camera observes the scene.
- Calibration between the projector and camera has to be known.
- Projection generated by a diffuser and difractive element of IR light,



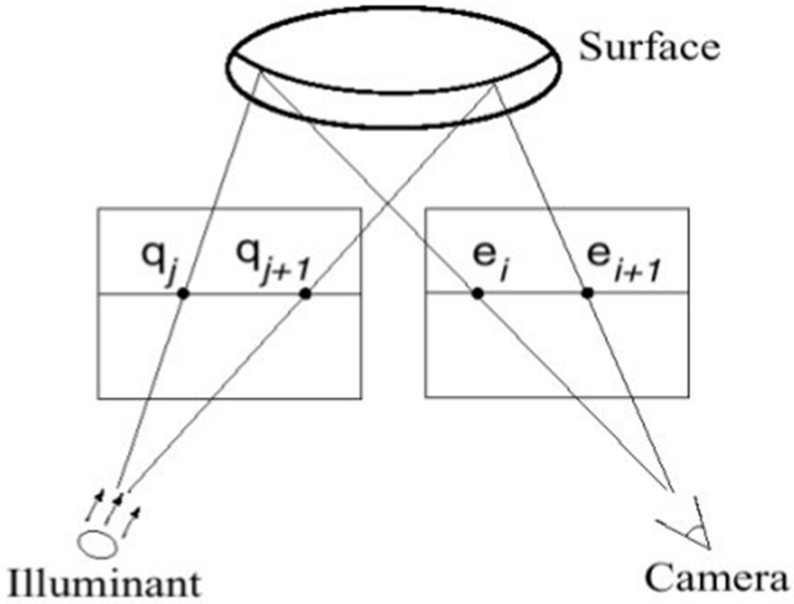
Picture from Primesence patent

Principles of Kinect (Primesensor)

How calculate the depth data?

- Triangulation of each speckle between a virtual image (pattern) and observed pattern.

- Each point has its correspondence speckle.



Picture from Primesence patent

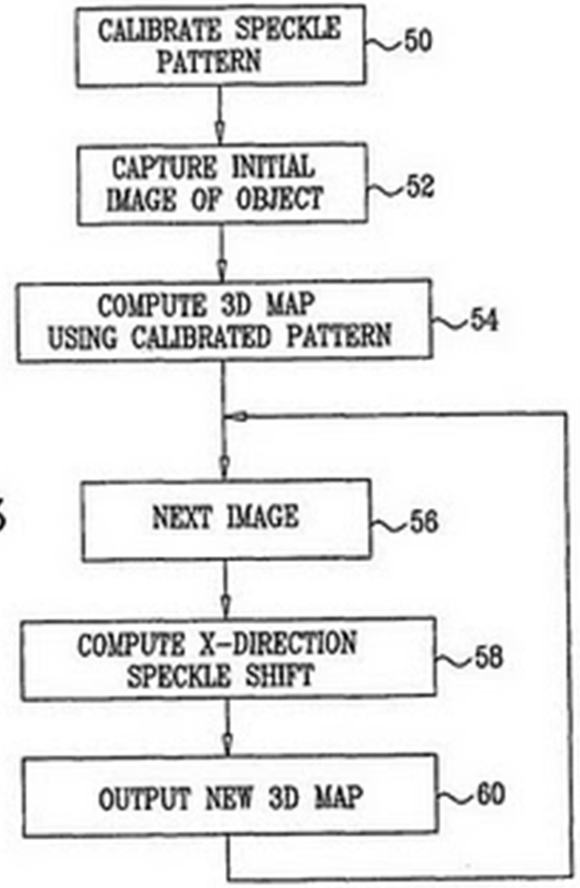
Principles of Kinect (Primesensor)

How calculate the depth data?

- Having a calibrated speckle pattern:
  - Compute the 3D map of the beginning frame.
  - Compute the x-direction speckle shift to renew the 3D map.

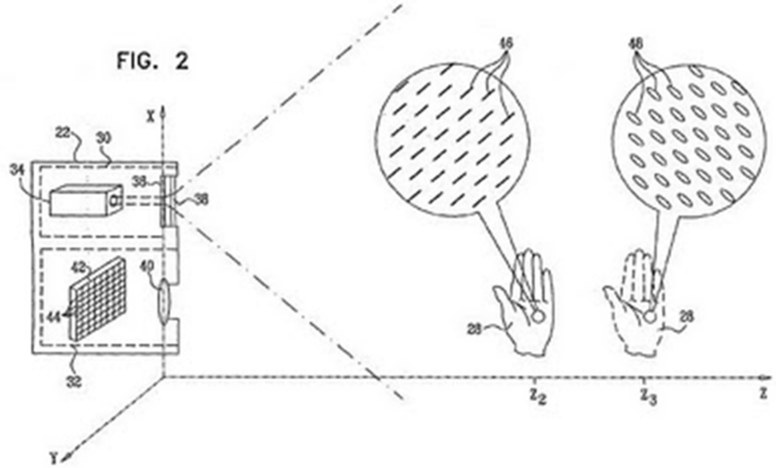
- Calibration is carried out the time of manufacture. A set of reference images were taken at different locations then stored in the memory. For the first computation.

FIG. 3



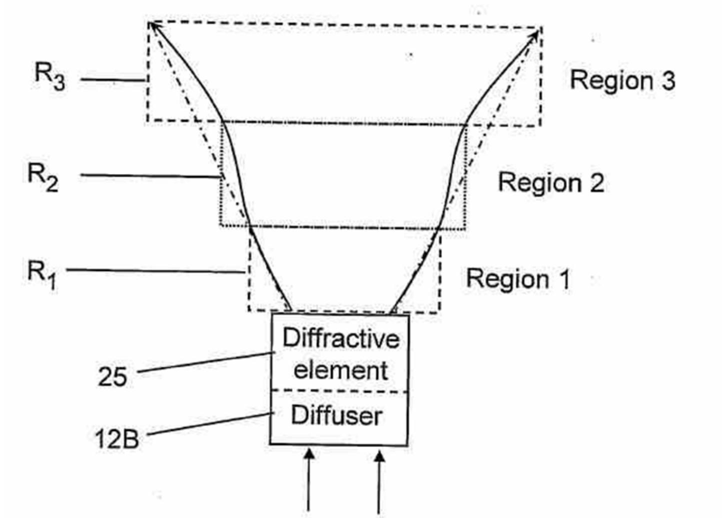
Picture from Primesence patent

Principles of Kinect (Primesensor)



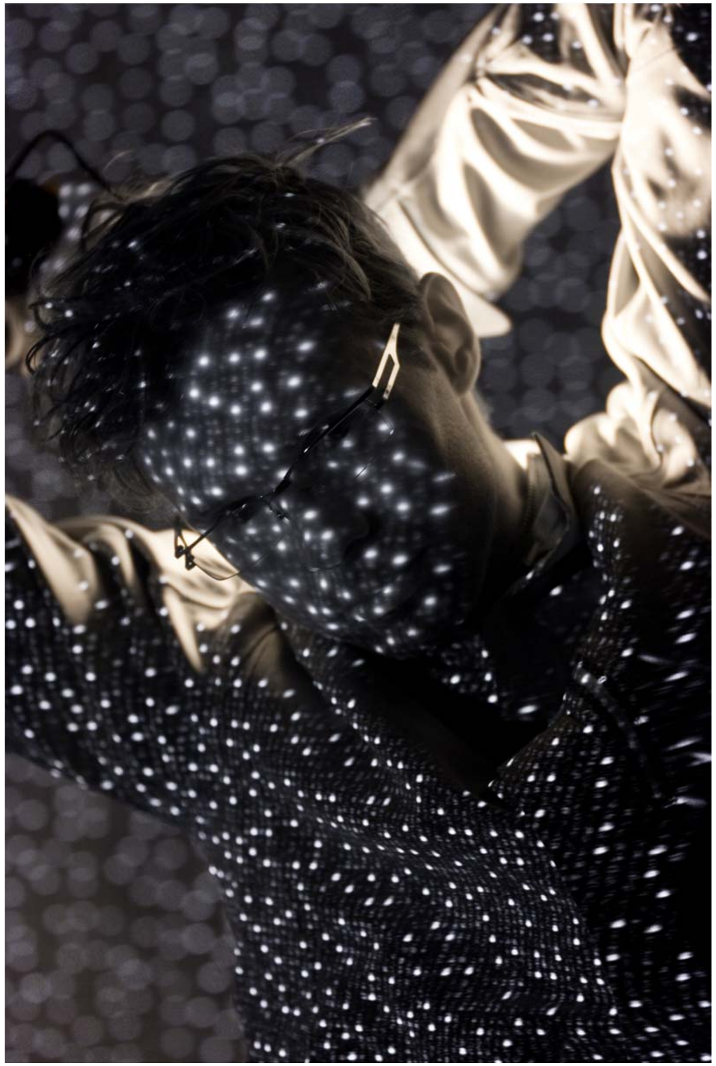
How Kinect works?

- The speckles size and shape depends on distance and orientation w.r.t. sensor.
- Kinect uses 3 different sizes of speckles for 3 different regions of distances.
- Then:
  - Near → High Accuracy
  - Far → Low accuracy



Picture from Primesence patent

Principles of Kinect (Primesensor)

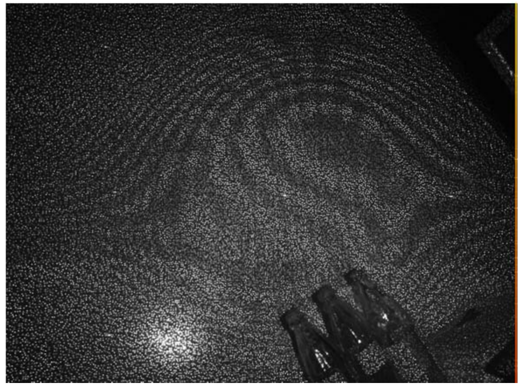


How pattern looks like?

- First Region: Allows to obtain a high accurate depth surface for near objects aprox. (0.8 – 1.2 m)
- Second Region: Allows to obtain medium accurate depth surface aprox. (1.2 – 2.0 m).
- Third Region: Allows to obtain a low accurate depth surface in far objects aprox. (2.0 – 3.5 m).

Picture from the Artist Audrey Penven

Principles of Kinect (Primesensor)

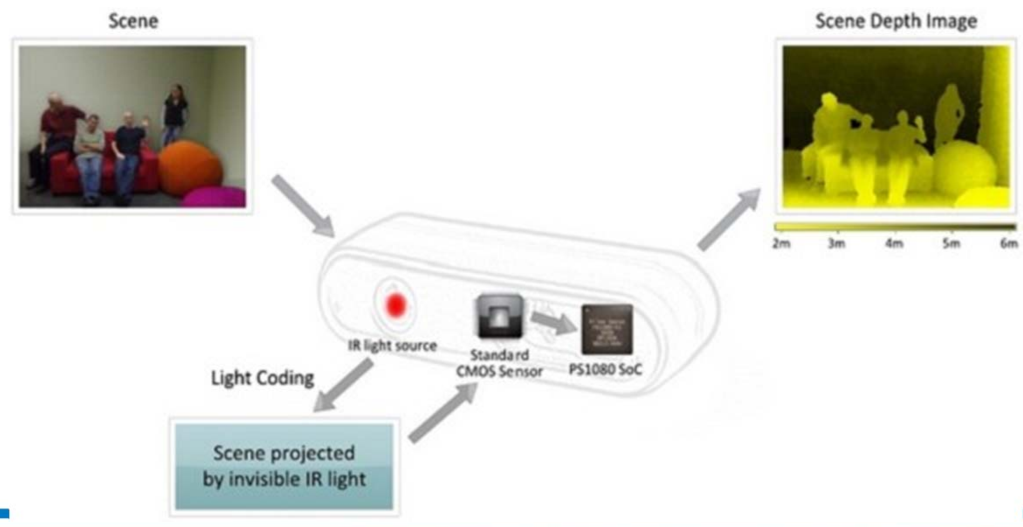


Projected Structured Light Pattern

Microsoft Kinect

- Depth resolution: 640x480 px
- RGB resolution: 1600x1200 px
- 60 FPS
- Operation range: 0.8m~3.5m
- spatial x/y resolution: 3mm @2m distance
- depth z resolution: 1cm @2m distance

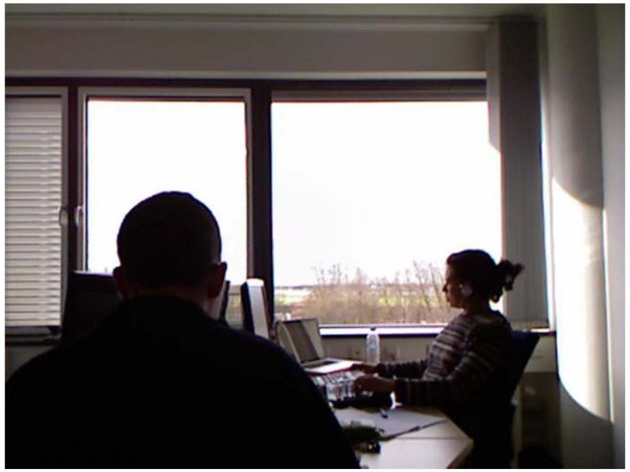
Picture from [15]



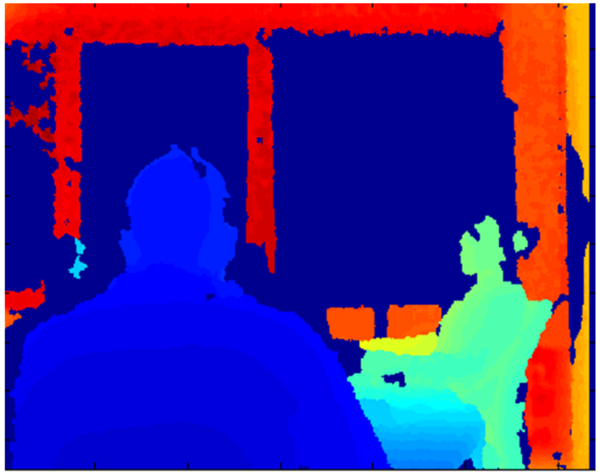
Other Range Imaging Techniques

Structured Light Imaging

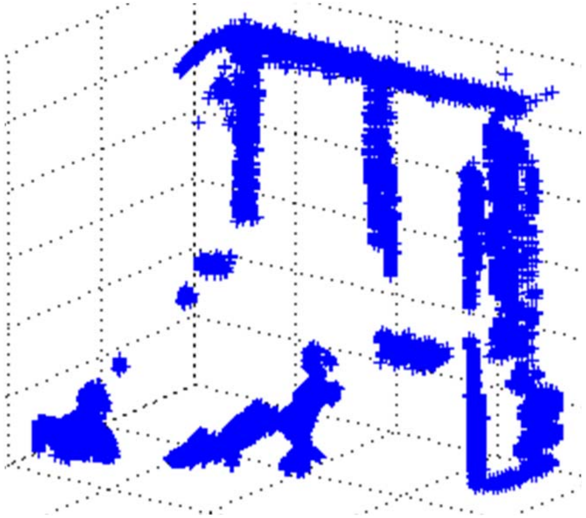
- Example: **Microsoft Kinect**



RGB Image



Depth Image



3D Reconstruction

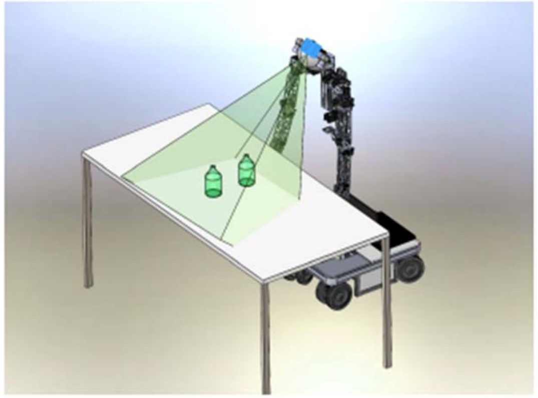
Lecture Outline

1. Introduction and Motivation
2. Principles of ToF Imaging
3. Computer Vision with ToF Cameras
4. Case Studies
5. Case Studies

Case Studies

Semantic Scene Analysis [4]

- Extract geometric representations from 3D point cloud data for object recognition
- Application: scene understanding for mobile robot
- RANSAC for fitting geometric models (e.g. plane, cylinders) to point data
- Points belonging to a detected model (e.g. table) are subsequently removed
- Final step: classification of remaining point clouds to object types



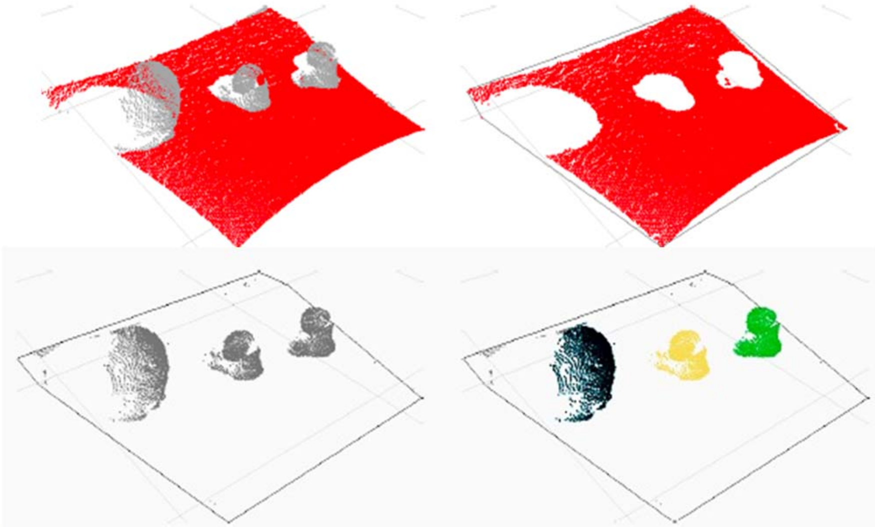


Image from [4]

Case Studies

Mixed/Augmented Reality [5]

- Real-time 3D scene augmentation with virtual objects
- Substitution for traditional chroma-keying (blue or green background) used in TV studios
- Combined ToF-RGB camera system
- Segmentation of moving objects
- Occlusions and shadows between real and virtual objects
- Tracking of camera location by co-registration of 3D depth data

Case Studies

Acquisition of 3D Scene Geometry [6]

- Combined ToF and RGB cameras
- Real-time acquisition of 3D scene geometry
- Each new frame is aligned to already previously aligned frames such that:
  - 3D geometry is matched
  - color information is matched
- Point cloud matching algorithm similar to Iterative Closest Points (ICP)
- Color information compensates for low depth image resolution
- Depth image compensates for hardly textured image regions

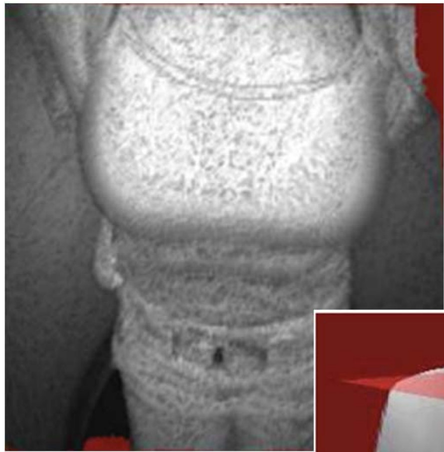
Case Studies

Simultaneous Localization and Mapping (SLAM) [7]

Case Studies

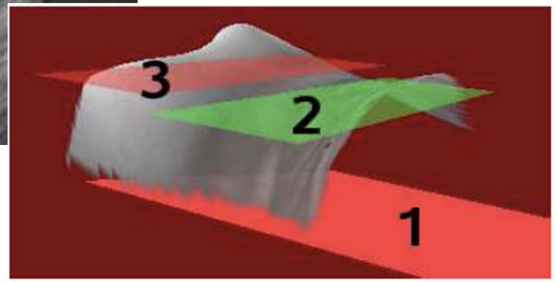
Medical Respiratory Motion Detection [8]

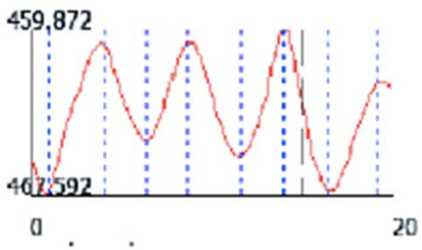
- Patient motion during examinations such as PET, CT causes artifacts
- Several breathing cycles during image acquisition
- Reduce artfacts when breathing motion pattern is known
- Measure breathing motion using ToF camera above patient
- Plane fitting to 3D data in specific regions of interest
- Continuous breathing signal



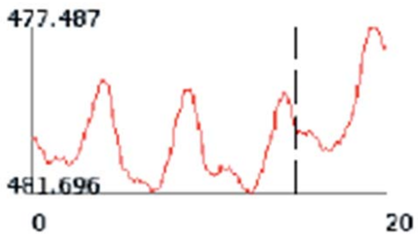
ToF data

Planes fitted





Chest motion

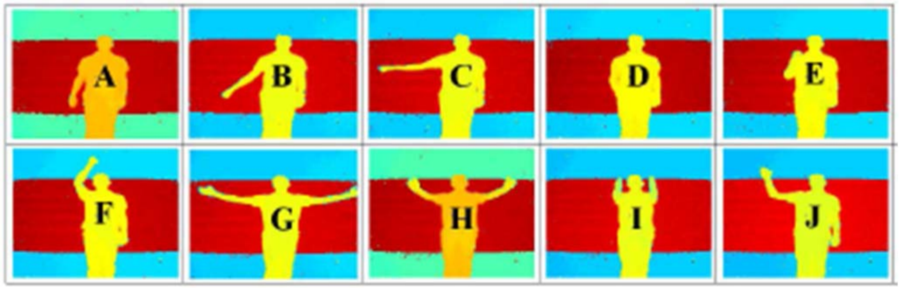


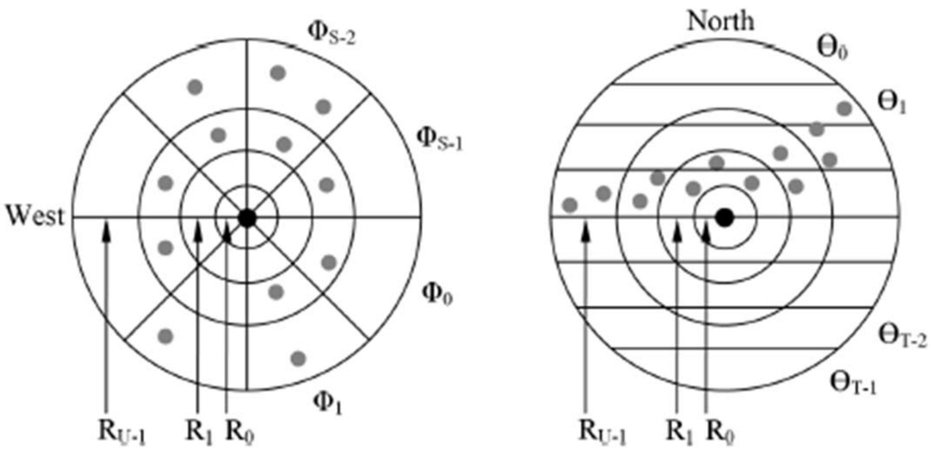
Abdomen motion

Case Studies

Gesture Recognition [9]

- Recognition of upper-body gestures
- Invariance to view-point changes (limited invariance)
- Representation of human point cloud using 3D shape context descriptors
- Rotational invariance by means of spherical harmonics functions

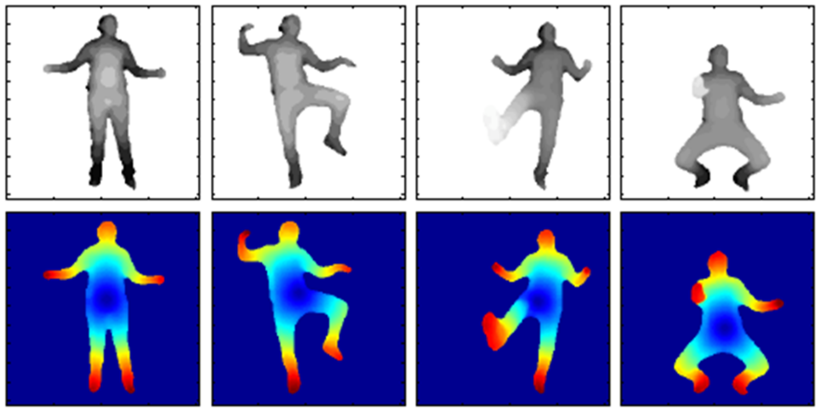


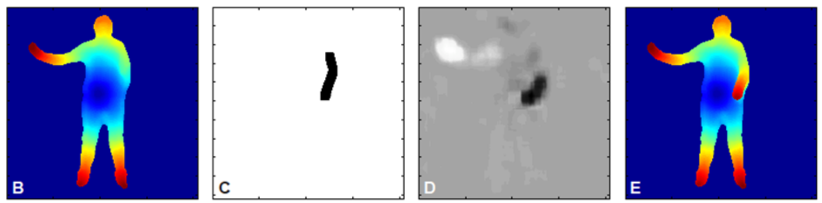


Case Studies

Markerless Human Motion Tracking [12]

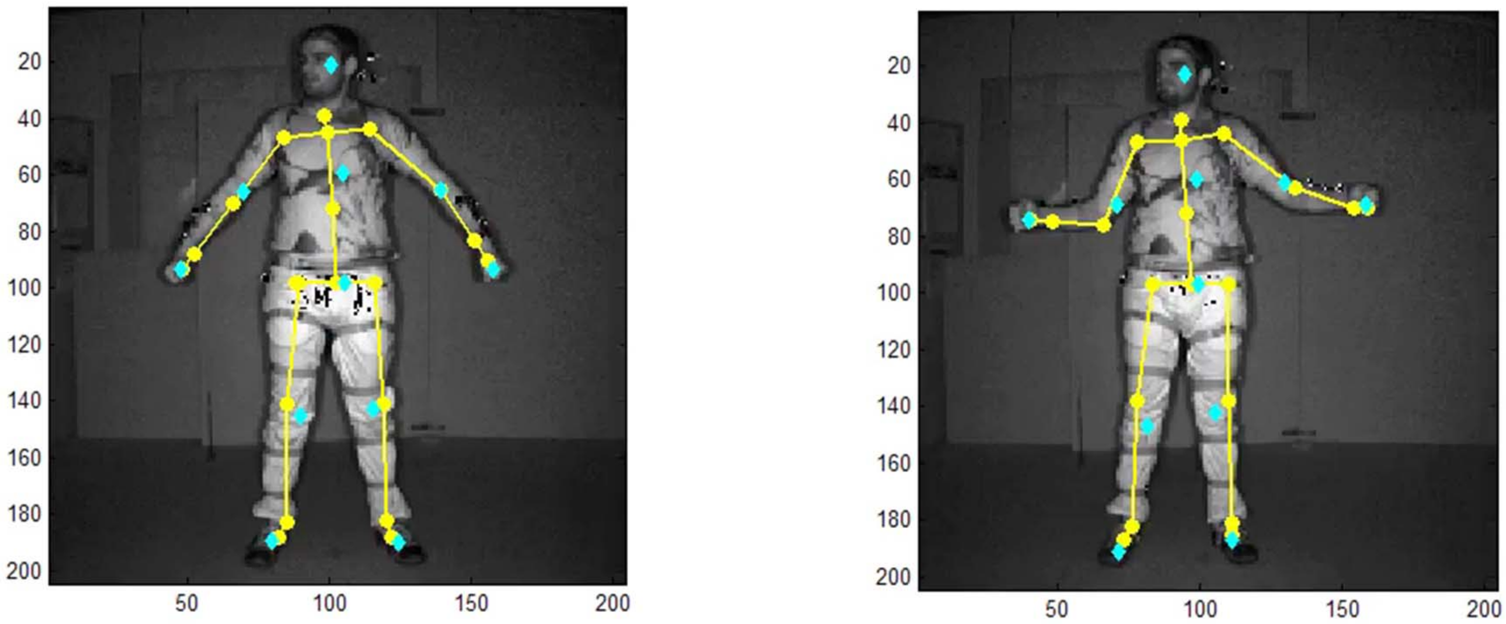
- Person segmentation by background subtraction
- Graph-based representation of 3D points
- Geodesic distance measurements (almost) invariant to pose changes
- Detection of anatomical landmarks as points with maximal geodesic distance from body center of mass
- Self-occlusion handling by means of motion information between frames
- Fitting skeleton to landmarks using inverse kinematics





Case Studies

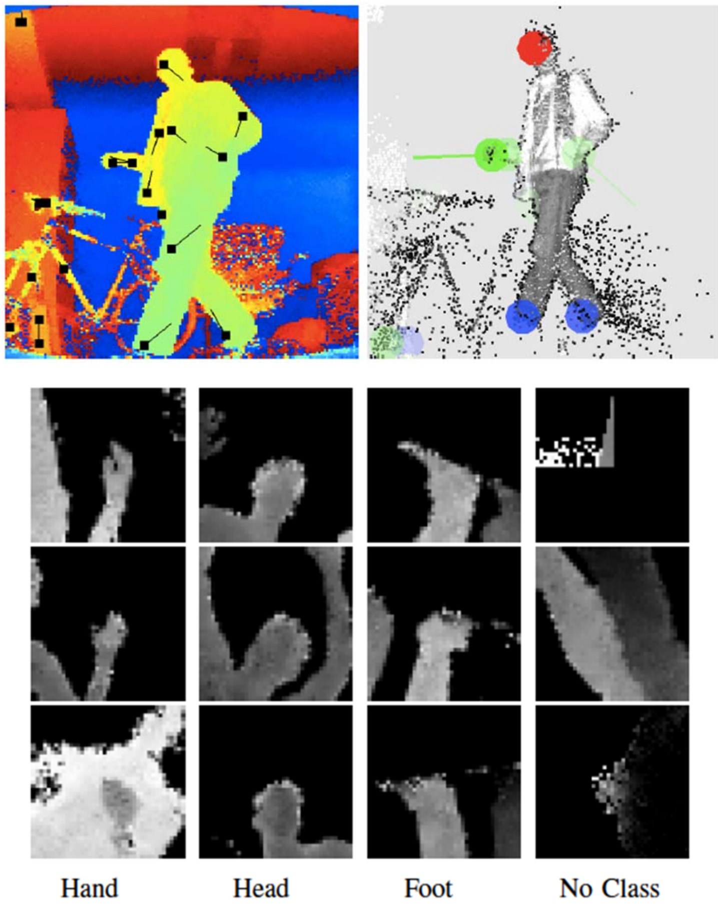
Markerless Human Motion Tracking [12]



Case Studies

Markerless Human Motion Tracking [10,11]

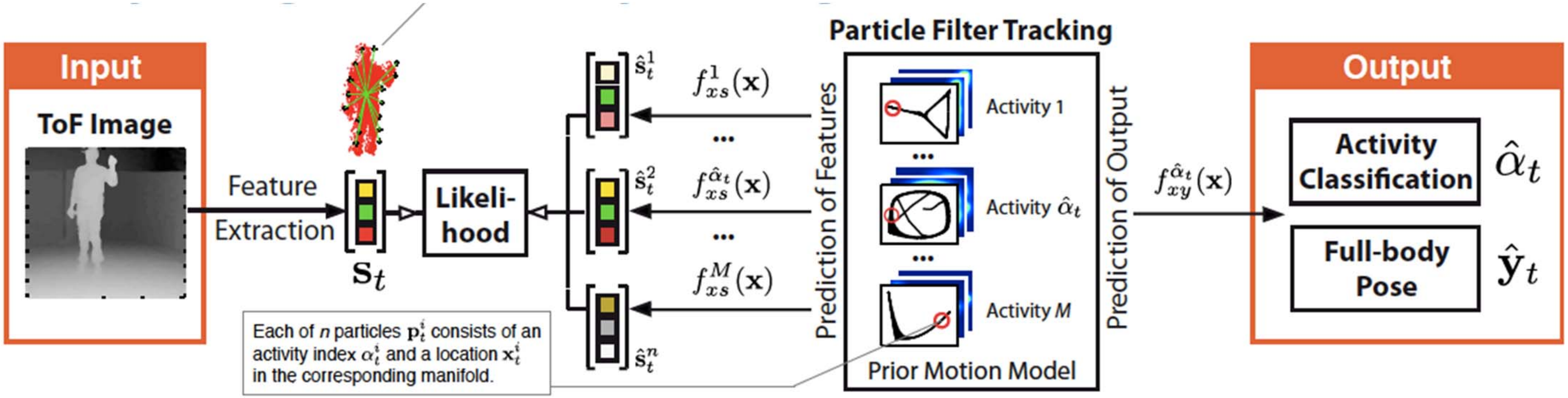
- Background segmentation
- Extraction of many interest points at local geodesic extrema with respect to the body centroid
- Classification as anatomical landmarks (e.g. head, hands, feet) using classifier trained on depth image patches



Case Studies

Human Body Tracking and Activity Recognition [13]

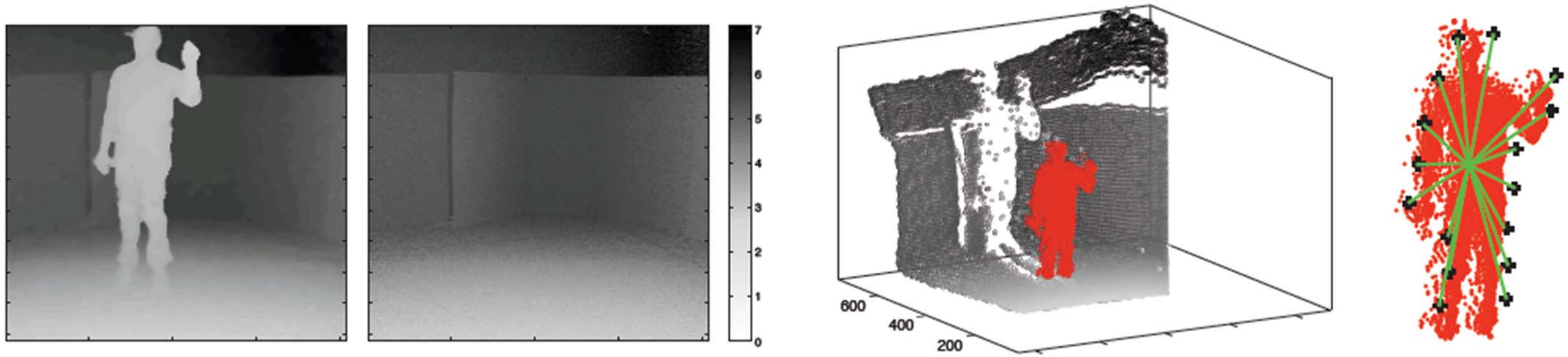
- Generative model with low-dimensional state space learned from training data
- Multiple-hypothesis tracking using particle filter
- Weighting of hyptheses by predicting ToF measurements and comparing to actual, true observations



Case Studies

Human Body Tracking and Activity Recognition [13]

- ToF-based feature descriptor for human poses
- Sampling of extremal points of 3D surface corresponding to person
- Features: distances of extremal points to centroid of point cloud
- Descriptor varies smoothly with motion



Case Studies

Manifold Learning for ToF-based
Human Body Tracking and Activity Recognition

BMVC 2010

Loren Arthur Schwarz, Diana Mateus, Victor Castaneda, Nassir Navab

Chair for Computer Aided Medical Procedures and Augmented Reality
Technische Universität München, Germany

References

[1] A. Kolb, E. Barth, R. Koch, R. Larsen: Time-of-flight sensors in computer graphics. Eurographics, 2009

[2] R. Lange: 3D Time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. PhD thesis, University of Siegen, 2000

[3] PMD Technologies GmbH, Siegen. http://www.pmdtec.com/

[4] D. Holz, R. Schnabel, D. Droeschel, J. Stückler, S. Behnke: Towards semantic scene analysis with Time-of-flight cameras. RoboCup Intermational Symposium, 2010

[5] R. Koch, I. Schiller, B. Bartczak, F. Kellner, K. Köser: MixIn3D: 3D mixed reality with ToF-Camera. Dynamic 3D Imaging, 2010

[6] B. Huhle, P. Jenke, W. Straßer: On-the-fly scene acquisition with a handy multi-sensor system. International Journal of Intelligent Systems Technologies and Applications, 2008

[7] V. Castañeda, D. Mateus, N. Navab: SLAM combining ToF and high-resolution cameras. IEEE Workshop on Motion and Video Computing, 2011

[8] J. Penne, C. Schaller, J. Hornegger, T. Kuwert: Robust real-time 3D respiratory motion detection using Time-of-flight cameras. International Journal of Computer Assisted Radiology and Surgery, 2008

References

[9] M. B. Holte, T. B. Moeslund, P. Fihl: Fusion of range and intensity information for view-invariant gesture recognition. Computer Vision and Pattern Recognition Workshops, 2008

[10] V. Ganapathi, C. Plagemann, D. Koller, S. Thrun: Real-time motion capture using a single Time-of-flight camera. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010

[11] C. Plagemann, V. Ganapathi, D. Koller: Real-time identification and localization of body parts from depth images. IEEE International Conference on Robotics and Automation (ICRA), 2010

[12] L. Schwarz, A. Mkhitaryan, D. Mateus, N. Navab: Estimating human 3D pose from Time-of-flight images based on geodesic distances and optical flow. IEEE Conference on Automatic Face and Gesture Recognition (FG), 2011

[13] L. Schwarz, D. Mateus, N. Navab: Manifold learning for ToF-based human body tracking and activity recognition. British Machine Vision Conference (BMVC), 2010

[14] E. Kollorz, J. Penne, J. Hornegger, A. Barke: Gesture recognition with a Time-of-Flight camera. International Journal of Intelligent Systems Technologies and Applications, 2008

[15] Microsoft Kinect. http://www.xbox.com/de-de/kinect