ImProc
Digital Image Processing

https://my.eurecom.fr/jcms/p0_2027226/en/improc

jean-luc.dugelay@eurecom.fr

ImProc. Digital Image Processing

https://my.eurecom.fr/jcms/p0_2027226/en/improc

6 Nov. 2018

Introduction to 3D Processing

Keyword 3D?

- Computer Graphics

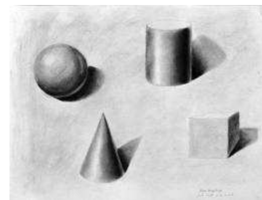
To model, to create, to manipulate, to animate 3D objects;
Realistic rendering of a synthetic scene (e.g. Ray tracing, Z-buffer)

- Signal/Image Processing

To extract 3D information of a scene from a (multi view) video

- Structure from Motion
- **Structure from Stereo**
- Shape from Shading
- Structure from Zooming

To customize some algorithms for 3D data (e.g. compression)

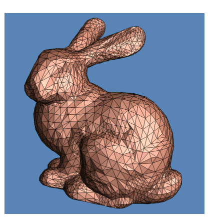


- TV & Movies

Stereoscopic visualization *of real or virtual scenes*

3D data acquisition (2.5D & 3D)

- From scratch (3D Mesh + texture)
- Using specific sensors (active or passive):
  - Multi-view acquisitions (including stereoscopic)
  - 3D Scanners like Minolta or Cyberware



Depth cameras (RGB-D)

Stereo Setup
- Passive system: no light is projected in the scene
- Two calibrated cameras
- Find points match between two images and triangulate to estimate distance
- Not very precise
- Not reliable in absence of textured surfaces

Time of Flight
- Measure light pulses round-trip from emitter to object to sensor
- Resolution depends on time measure precision
- ~30 fps
- Indoor and outdoor

Structured light
- Measure distortion in a pattern of IR light and shadow projected over the scene
- Very fast and precise
- Only indoor

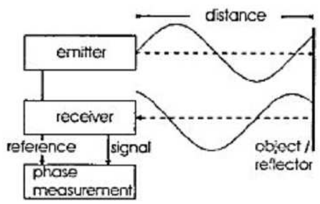
Light Field
- Passive system
- Find the direction where the light is from
- Sensible to light variation
- Indoor and outdoor

What is a ToF camera?

- Provides intensity data and range information for every pixel simultaneously
- Equipped with one IR emitter and one detector
  - *Two main estimation techniques:*
    - *Pulsed Modulation*
    - *Continuous Wave Modulation*

Optex ZC1000

- Estimates the distance computing the time needed for a light pulse to travel to the target and return to the camera after being scattered ($c = 3 \cdot 10^8$m/s)
- Advantages:
  - Direct measurement of time of flight
  - Limited influence of background illumination
  - Emitter and detector directions are collinear
- Disadvantages:
  - High accuracy required
  - Error due to light scattering
  - Pulses generation is hard to achieve



ToF - Principles

Continuous Wave Modulation

- Advantages:
  - Various light sources can be used
  - Different modulation techniques available
  - Simultaneous range and amplitude images
- Disadvantages:
  - Integration time:
    - Low frame-rate
    - Motion blur

Kinect

Projects a pseudo-random IR light pattern (Speckles) into the scene

The CMOS IR camera records the pattern distortion (x-direction shift)

For the first frame

Compute the triangulation of each speckle between the two patterns (virtual and observed) creating the 3D map

For the following frames

Re-compute the x-direction speckle shift and updates the 3D map

Pattern:

Three different sizes of speckles

Three different regions of distances FIRST (0.8 – 1.2 m) SECOND (1.2 – 2 m) THIRD (2 – 3.5 m)

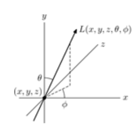
ToF vs Kinect

Strengths & weaknesses

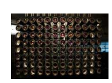
Feature	ToF	Kinect
Illumination condition	Outdoor / indoor	Indoor
Depth range	10 m (extendable up to 30m)	3.5 m
Depth accuracy	mm	cm
Resolution	Up to 200x200 pixels	640x480 pixels
Measurement technique	Pulsed / Continuous -Wave Modulation	Triangulation
Computational efforts	On-board FPGA for phase and intensity measurement	Very high costly computations (parallel processor)
Market	Industrial	Video Games
Price	From 3000€	150€

Light Field history

- **1991**: Adelson and Bergen officialised plenoptic function $P(\theta, \phi, \lambda, t, x, y, z)$

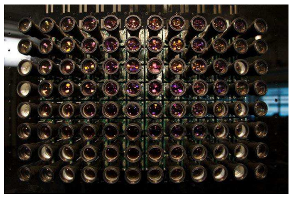


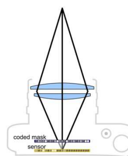
- **1996**: Levoy and Hanrahan published a paper about light field function $L(u, v, x, y)$
  - "extracting 2D slices from this 4D function one could construct new perspective views of a scene"
- **2005**: Wilburn (Stanford University) created a gantry cameras



- **2006**: Ng (Stanford University) published his PhD thesis and founded Lytro company
- **2010**: First commercial light field cameras from Lytro and Raytrix company

Light field cameras



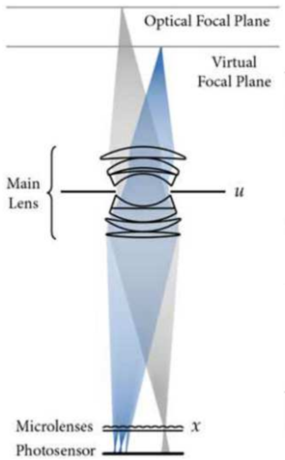


Comparison with other 3D technologies

3D measure technics	Time of Flight	Stereovision	Structured light	Light Field
XY resolution	Good	Variable	Average	Variable
Z precision	Average	Average	Really good	To be investigate
Performance in outdoor environment	Average	Good	Low	Good
Computation complexity	Low	High	High/Average	Low
N acquisition	1	2	Several/Video	1
Data format	Depth map + RGB	Depth map + RGB	3D vision	Depth map + Multiview images + EPI + refocused images
Cost	Average	Average	Variable	Variable

Courtesy of Professor Christophe CUDEL from U. Mulhouse

Lytro camera

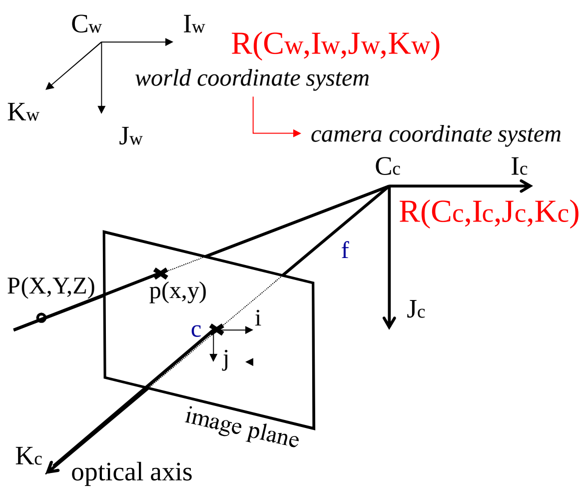


Light field images contain information not only about light intensity but also about the direction where it comes from.

It is possible to obtain a light field camera inserting a micro-lens array between the sensor and the main lens of a standard camera.

« depth of field » (DOF)
The area within the depth of field appears sharp, while the areas beyond the depth of field appear blurry.

Calibration: internal & external parameters



Internal
- ***f*** *focal length*
- ***c*** *projection of C (optical centre) in the image plane*

[by default c(0,0,f)]

External
- **T**ranslation
- **R**otation

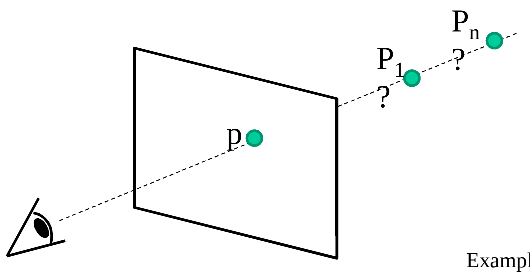
camera pose & orientation

Internal parameters

$$\begin{cases} x = f\,\dfrac{X}{Z} \\ y = f\,\dfrac{Y}{Z} \end{cases}$$

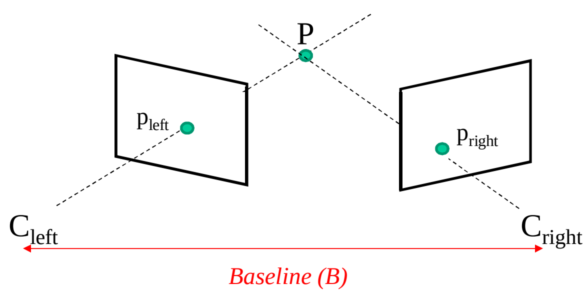
The knowledge of **internal parameters** yields:

- to associate to a point P of the scene its corresponding point p in the image plane
- or inversely, to associate to a point p, a half line including P



Example of specific patterns used for calibration

Structure from Stereo

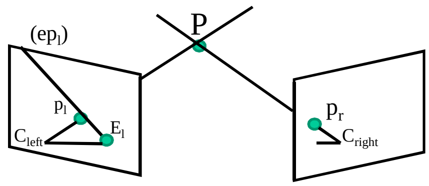


- Stereo vision systems take two images of a scene from 2 different viewpoints
- Disparity: Given a 3D point P, it is the difference of projected positions into the left (p_l) and right (p_l) images.
- From disparity, we can derive depth.

  If the two video cameras are parallel: $\begin{cases} x_r - x_l = K/Z \quad \text{with} \quad K = f.B \\ y_l = y_r \end{cases}$

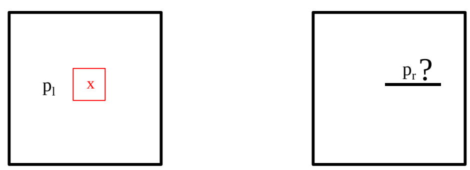
Epipolar geometry

disparity map



1-D search

E_l Intersection between line (C_l,C_r *)and left image plane;*
Ep_l *left epipolar (half) line associated with the right pixel p_r*



Specific case of parallel cameras

Structure from Motion

basic equations linking 2D & 3D Motion

Relative motion of an (rigid) object w.r.t. camera can be decomposed as:

- an instantaneous rotational motion (W)
- and an instantaneous translational motion (T)

That is to say, between two times t and t', we have:

$$P' = W.P + T$$

Under the following assumptions:

- Rotations are small;
- View angle relatively small;
- Translation along optical axis small compared to the distance object/camera;

(Normalization: f =1)

$$\begin{cases} v_x(x,y) = +y.w_z - (1+x^2).w_y + x.y.w_x + \frac{1}{Z}.t_x - \frac{x}{Z}.t_z \\ v_y(x,y) = -x.w_z - (1+y^2).w_x + x.y.w_y + \frac{1}{Z}.t_y - \frac{y}{Z}.t \end{cases}$$

P'=W.P+T
if rotations are small, W can be expressed as,

$$\begin{pmatrix} 1 & w_z & -w_y \\ -w_z & 1 & w_x \\ w_y & -w_x & 1 \end{pmatrix}$$

We then obtain,

$$\begin{cases} X' = X + w_z.Y + -w_y.Z + t_x \, (a) \\ Y' = -w_z.X + Y + w_x.Z + t_y \, (b) \\ Z' = w_y.X + -w_x.Y + Z + t_z \, (c) \end{cases}$$

Then,

$$\begin{cases} X'/Z' = \left(X + w_z.Y + -w_y.Z + t_x\right) / \left(w_y.X + -w_x.Y + Z + t_z\right) \quad (a/c) \\ Y'/Z' = \left(-w_z.X + Y + w_x.Z + t_y\right) / \left(w_y.X + -w_x.Y + Z + t_z\right) \quad (b/c) \end{cases}$$

by dividing both numerator and denominator by Z,

$$\begin{cases} X'/Z' = \left(X/Z + w_z.Y/Z + -w_y.Z/Z + t_x/Z\right) / \left(w_y.X/Z + -w_x.Y/Z + Z/Z + t_z/Z\right) \\ Y'/Z' = \left(-w_z.X/Z + Y/Z + w_x.Z/Z + t_y/Z\right) / \left(w_y.X/Z + -w_x.Y/Z + Z/Z + t_z/Z\right) \end{cases}$$

Assuming that f=1

$$\begin{cases} x = f\,\dfrac{X}{Z} \\ y = f\,\dfrac{Y}{Z} \end{cases}$$

$$\begin{cases} x' = \left(x + w_z.y + -w_y + t_x/Z\right) / \left(w_y.x + -w_x.y + 1 + t_z/Z\right) \\ y' = \left(-w_z.x + y + w_x + t_y/Z\right) / \left(w_y.x + -w_x.y + 1 + t_z/Z\right) \end{cases}$$

$$\begin{cases} v_x(x,y) = x' - x \\ v_y(x,y) = y' - y \end{cases}$$

$$\begin{cases} v_x = \left(w_z.y + -w_y(1+x^2) + w_x.x.y + t_x/Z - x.t_z/Z\right)/D \\ v_y = \left(-w_z.x + -w_x(1+y^2) - w_x.x.y + t_y/Z - y.t_z/Z\right)/D \end{cases}$$

$$D = w_y.x + -w_x.y + 1 + t_z/Z$$

-View angle relatively small (<<1);
-Translation along optical axis small compared to the distance object/camera;

$$\begin{cases} v_x(x,y) = +y.w_z - (1+x^2).w_y + x.y.w_x + \dfrac{1}{Z}.t_x - \dfrac{x}{Z}.t_z \\ v_y(x,y) = -x.w_z - (1+y^2).w_x + x.y.w_y + \dfrac{1}{Z}.t_y - \dfrac{y}{Z}.t \end{cases}$$

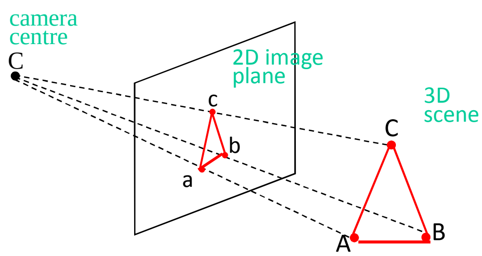
Planar facet

$$Z_{(x,y)} = A.X + B.Y + C$$

$$\frac{1}{Z_{(x,y)}} = k_x.x + k_y.y + k_y$$

Basic 2D/3D equations can be simplified as,

$$\begin{cases} v_x(x,y) = a_1 + a_2.x + a_3.y \\ v_y(x,y) = a_4 + a_5.x + a_6.y \end{cases}$$



$$\begin{cases} v_x(x,y) = a_1 + a_2.x + a_3.y + a_7.x.y + a_8.x^2 \\ v_y(x,y) = a_4 + a_5.x + a_6.y + a_8.x.y + a_7.y^2 \end{cases}$$

$$\begin{cases} a_1 = -w_y + k_z.t_x \\ a_2 = k_x.t_x - k_z.t_z \\ a_3 = w_z + k_y.t_x \\ a_4 = w_x + k_z.t_y \\ a_5 = -w_z + k_x.t_y \\ a_6 = k_y.t_y - k_z.t_z \\ a_7 = w_x - k_y.t_z \\ a_8 = -w_y - k_x.t_z \end{cases}$$

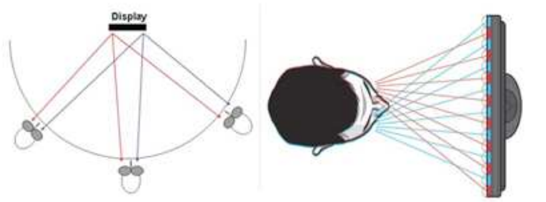
(auto) Stereoscopic Visualization

- Left eye must see images acquired by the left sensor,

and the right eye must see only images acquired by the right sensor.

Technologies with or without glasses (glasses-free 3DTV)

- Anaglyph (glasses where the two lenses are different red/cyan)
- Lenticular lens system



[Left] Watching glasses-type 3D TV
[Right] Watching single point, glasses-free 3D TV