Lecture 8

Basics of Computer Vision for Real-Time Estimating the Pose of the Ball of the Butterfly Robot

---

Maksim Surov
Stockholm, 2018-10-22

Learning outcomes: projective geometry, camera model, camera intrinsic calibration, perspective-n-point problem, camera extrinsic calibration, estimate position of a ball, estimate orientation of a ball

Outline

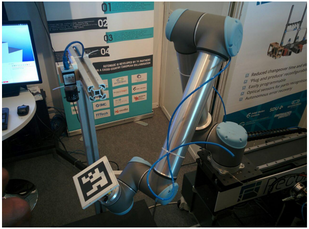
1. Computer Vision in Robotics: Examples
2. Camera Inside
3. Mathematica Model of Camera
4. Estimate Object Pose
5. Estimate Position of a Ball

Computer Vision in Robotics: Examples

Computer Vision Applications

Most of the robotic systems use camera sensors to estimate position, collisions, humans, etc

object detection (classification) problem



pose estimation

Object pose estimation

kuka.mp4

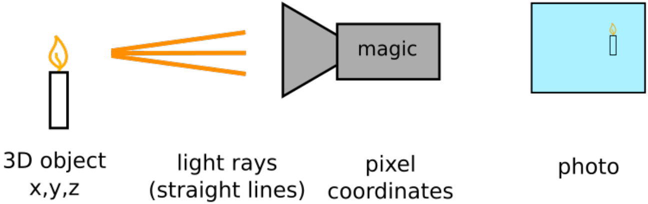
Object pose estimation

juggle.webm-00.00.48.592-00.01.17.635.webm

Camera Inside

Camera Schematic

Forward problem:

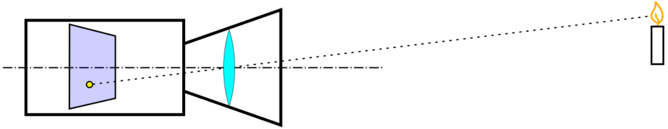


Our goal is the inverse problem:

pixel coordinates $\mapsto$ 3D coordinates.

Camera Schematic

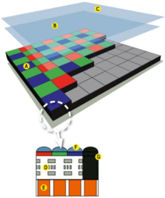
Each pixel measures "irradiance" in a selected direction



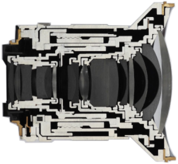
sensor
(grid of sensors)

lens
(system of lenses)

light
source in 3D



The sensor measures the irradiance in a given point.



The lens must select the rays such that each pixel receives light from one direction only.

A ray travels through lens and hits to the sensor.

Camera Lens

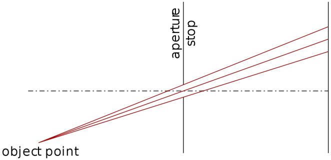
Main goal: split the rays!

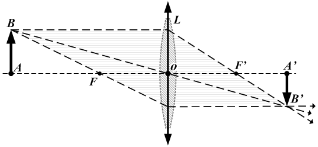
Pinhole optics compromise:

- sharp
- bright

In a dark room there is a point source of light, all the pixels must be zero except one, i.e. projection of a point is also a point

Geometrical optics, focal length, angle of view





Camera Sensor

Pixel value is equivalent to "number" of photons received by the sensor's pixel

$u = f(E \cdot \tau)$, where

$f : \mathbb{R}^+ \to [0, 255]$	camera response function
$\tau \in \mathbb{R}^+$	exposure time
$E \in \mathbb{R}^+$	irradiance

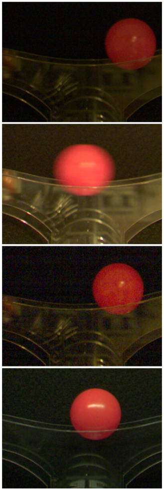
The pixels value depends on

- irradiance
- exposure time
- camera gain, gamma, etc

Sensor Parameters

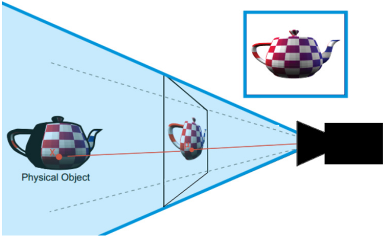
- exposure
- gain
- aperture
- resolution
- SNR
- shutter-type (global,roller)
- sensor type (CCD,CMOS)

exposure	gain	light
20ms	0db	off
100ms	0db	off
10ms	12db	off
1ms	0db	on



Projection

- forward map: 3d object → image
- inverse problem: image → 3d object coordinates

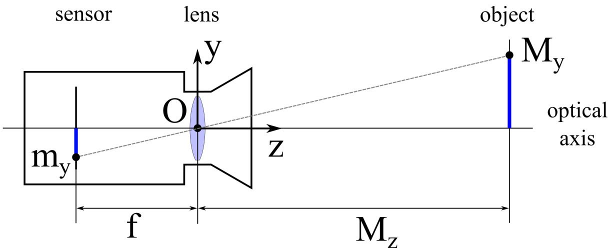


How to find the forward map?

Mathematica Model of Camera

Pinhole Camera

No refraction:



Similar triangles:

$$\frac{m_y}{f} = \frac{M_y}{M_z}$$
$$\frac{m_x}{f} = \frac{M_x}{M_z}$$

The same:

$$\begin{pmatrix} m_x \\ m_y \\ f \end{pmatrix} = \lambda \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix}$$

with

$$\lambda = \frac{f}{M_z}$$

Equivalence operator

The equivalence operator

$$\begin{pmatrix} m_x \\ m_y \\ f \end{pmatrix} = \lambda \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} \iff \begin{pmatrix} m_x \\ m_y \\ f \end{pmatrix} \sim \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix}$$

The tuples

$$a = (a_x, a_y, a_z),$$
$$b = (b_x, b_y, b_z) \in \mathbb{R}^3$$

are called equivalent

$$a \sim b \iff \exists \lambda \in \mathbb{R} \setminus \{0\} : a = \lambda \cdot b$$

Examples:

- $a \sim a \quad \forall a \in \mathbb{R}^3$
- $a \sim 2 \cdot a \quad \forall a \in \mathbb{R}^3$
- if $a \sim b$ then $K \cdot a \sim K \cdot b \quad \forall K \in \mathbb{R}^{3\times3}$

Equivalence operator

The tuples

$$a = (a_x, a_y, a_z),$$
$$b = (b_x, b_y, b_z) \in \mathbb{R}^3$$

are called equivalent

$$a \sim b \iff \exists \lambda \in \mathbb{R} \setminus \{0\} : a = \lambda \cdot b$$

Examples:

- $a \sim a \quad \forall a \in \mathbb{R}^3$
- $a \sim 2 \cdot a \quad \forall a \in \mathbb{R}^3$
- if $a \sim b$ then $K \cdot a \sim K \cdot b \quad \forall K \in \mathbb{R}^{3 \times 3}$

Equivalence operator

The tuples

$$a = (a_x, a_y, a_z),$$
$$b = (b_x, b_y, b_z) \in \mathbb{R}^3$$

are called equivalent

$$a \sim b \iff \exists \lambda \in \mathbb{R} \setminus \{0\} : a = \lambda \cdot b$$

Examples:

- $a \sim a \quad \forall a \in \mathbb{R}^3$
- $a \sim 2 \cdot a \quad \forall a \in \mathbb{R}^3$
- if $a \sim b$ then $K \cdot a \sim K \cdot b \quad \forall K \in \mathbb{R}^{3 \times 3}$

Equivalence operator

The tuples

$$a = (a_x, a_y, a_z),$$
$$b = (b_x, b_y, b_z) \in \mathbb{R}^3$$

are called equivalent

$$a \sim b \iff \exists \lambda \in \mathbb{R} \setminus \{0\} : a = \lambda \cdot b$$

Examples:

- $a \sim a \quad \forall a \in \mathbb{R}^3$
- $a \sim 2 \cdot a \quad \forall a \in \mathbb{R}^3$
- if $a \sim b$ then $K \cdot a \sim K \cdot b \quad \forall K \in \mathbb{R}^{3 \times 3}$

Geometrical Meaning

There is a straight line given in parametric form

$$p(t) = a \cdot t$$

where $a = (a_x, a_y, a_z)$

The line can be re-parametrized by $\tau = \frac{t}{\lambda}$ for some λ, then

$$p(t) = a \cdot \lambda\tau(t) = b \cdot \tau(t)$$

These are the same straight line with different affine parameterizations!

$$a \sim b = \lambda a$$

the same straight line $\iff$ the tuples are equivalent

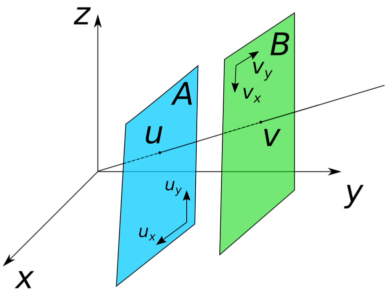
Geometrical Meaning: Homography

Given 2 planes in parametric form

$$(x, y, z)^T = A \cdot (u_x, u_y, 1)^T$$
$$(x, y, z)^T = B \cdot (v_x, v_y, 1)^T$$

A straight line crossing them

$$(x, y, z) = a \cdot t$$



Intersection points: $a \sim A \cdot u, \quad a \sim B \cdot v$

then $\quad A \cdot u \sim B \cdot v$

then

$$u \sim A^{-1} B \cdot v = H \cdot v$$

H is homography matrix

Geometrical Meaning: Homography

Let we have n pairs of points

$$u^i \sim H \cdot v^i$$

and we need to find H

$$\left(u_x^i, u_y^i, 1\right)^T = \lambda^i \cdot H \cdot \left(v_x^i, v_y^i, 1\right)^T$$

The last row gives $\lambda^i = \frac{1}{\mathrm{row}_3(H) \cdot v^i}$. Then multiplying by λ^i we have

$$u_x^i \cdot \mathrm{row}_3\left(H\right) \cdot v^i = \mathrm{row}_1\left(H\right) \cdot v^i$$
$$u_y^i \cdot \mathrm{row}_3\left(H\right) \cdot v^i = \mathrm{row}_2\left(H\right) \cdot v^i$$

This is a system of linear eqs wrt H! One pair gives 2 equations.
$\Rightarrow$ we need 4 pairs to find H!

Pinhole Camera Model

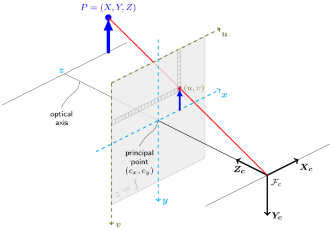
Projection of a point-size light source at have the coordinates

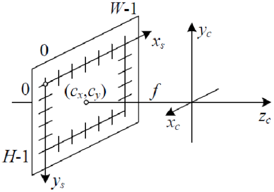
$$(m_x, m_y, f) \sim (M_x, M_y, M_z)$$

meters → pixels

$$u_x = \frac{m_x}{d_x} + c_x$$
$$u_y = \frac{m_y}{d_y} + c_y$$

d_x, d_y	pixel width and height [m]
c_x, c_y	shift of sensor relative to main optical axis [px]





Pinhole Camera

The same transformation in matrix form

$$\begin{pmatrix} u_x \\ u_y \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{d_x} & 0 & \frac{c_x}{f} \\ 0 & \frac{1}{d_y} & \frac{c_y}{f} \\ 0 & 0 & \frac{1}{f} \end{pmatrix} \begin{pmatrix} m_x \\ m_y \\ f \end{pmatrix}$$

multiply by f

$$\begin{pmatrix} u_x \\ u_y \\ 1 \end{pmatrix} \sim \underbrace{\begin{pmatrix} \frac{f}{d_x} & 0 & c_x \\ 0 & \frac{f}{d_y} & c_y \\ 0 & 0 & 1 \end{pmatrix}}_{=:K} \begin{pmatrix} m_x \\ m_y \\ f \end{pmatrix}$$

K is a camera intrinsics matrix.

Due to $(m_x, m_y, f) \sim M$:

$$u \sim K \cdot M$$

Equation of Projection

The projection equation is given by

$$u \sim K \cdot M$$

where

$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

$u = (u_x, u_y, 1)$	projection coordinates [px]
$M = (M_x, M_y, M_z)$	3D point coordinates in camera CS frame! [m]
$f_x = \frac{f}{d_x}, f_y = \frac{f}{d_y}$	focal distances [px]
c_x, c_y	principal point [px]

Geometrical Meaning

There is a ray (straight line)

$$a = (a_x, a_y, a_z)$$

Camera matrix transforms the ray into the coordinates of projection

$$K : a \mapsto u$$

$$K : \text{ray of light} \mapsto \text{pixel coordinates}$$

The inverse operator

$$K^{-1} : \text{pixel coordinates} \mapsto \text{ray of light}$$

Nonlinear Optical Effects

The lens is not perfect!

- lens is not thin $\rightarrow$ **radial distortion**
- lenses are not exactly parallel $\rightarrow$ **tangential distortion**

1. the straight line $(x, y, 1) \sim (M_x, M_y, M_z)$
2. nonlinear distortion

$$x' = x\left(1 + k_1 r^2 + k_2 r^4 + ...\right) + 2p_1 xy + p_2\left(r^2 + 2x^2\right)$$
$$y' = \underbrace{y\left(1 + k_1 r^2 + k_2 r^4 + ...\right)}_{\text{radial}} + \underbrace{p_1\left(r^2 + 2y^2\right) + 2p_2 xy}_{\text{tangential}}$$

3. geometrical transformation $u_x = f_x \cdot x' + c_x$, $u_y = f_y \cdot y' + c_y$

Geometrical Meaning

There is a ray

$$a = (a_x, a_y, a_z)$$

the ray is (nonlinearly) distorted by the lens

$$\text{distortion} : \ a \mapsto a'$$
$$K : \ a' \mapsto u$$

the inverse operator

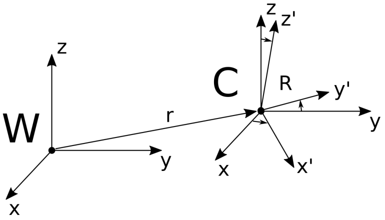
$$K^{-1} u \mapsto a'$$
$$\text{undistortion} : \ a' \mapsto a$$

we can reconstruct the rays!

Camera Extrinsics

Camera position relative to the world coordinate system frame is given by rotation R_{cw}, and displacement r_{cw}:

$$M_c = R_{cw} \cdot M_w + r_{cw}$$



the final projection law is

$$u \sim K \cdot \begin{pmatrix} I & 0 \end{pmatrix} \cdot T_{cw} \cdot \begin{pmatrix} M_w \\ 1 \end{pmatrix} = K \cdot \begin{pmatrix} R_{cw} & r_{cw} \end{pmatrix} \cdot \begin{pmatrix} M_w \\ 1 \end{pmatrix}$$

homogeneous transformation matrix = camera extrinsics

$$T_{cw} = \begin{pmatrix} R_{cw} & r_{cw} \\ 0 & 1 \end{pmatrix}$$

Camera Parameters

1. intrinsics: f_x, f_y, c_x, c_y
2. tangential distortion: k_1, k_2, k_3
3. radial distortion: p_1, p_2
4. extrinsics: $r \in \mathbb{R}^3, R \in SO(3)$

intrinsics:	$4 + 3 + 2 = 9$ values
extrinsics:	6 values
total:	$4 + 3 + 2 + 6 = 15$ values!

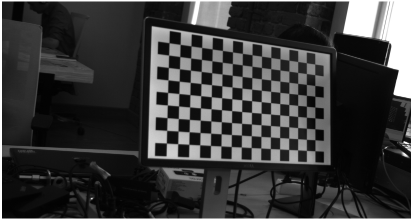
How to find all of them?

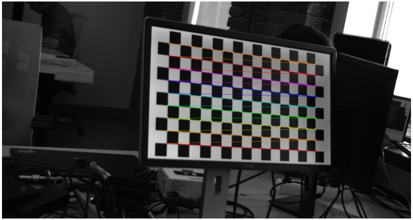
Camera Intrinsics Calibration

Find f_x, f_y, c_x, c_y:

1. forget about distortion
2. corners have the coords $v^i = (v^i_x, v^i_y, 1)$
3. take photo of pattern
4. detect corners on picture: u^i
5. find homography: $u^i \sim H \cdot v^i$ (need at least 4 corners)
6. decompose homography onto extrinsics and intrinsics: $H \sim K \cdot \begin{pmatrix} R_{cb} & r_{cb} \end{pmatrix}$

homography: 8 indep components
R_{cb}, r_{cb}: 6 indep components
$\Rightarrow$ + 2 excessive equations!
$\Rightarrow$ we need at least 2 photos to find 4 parameters!





Camera Intrinsics Calibration

Find k_1, k_2, k_3, p_1, p_2 iteratively:

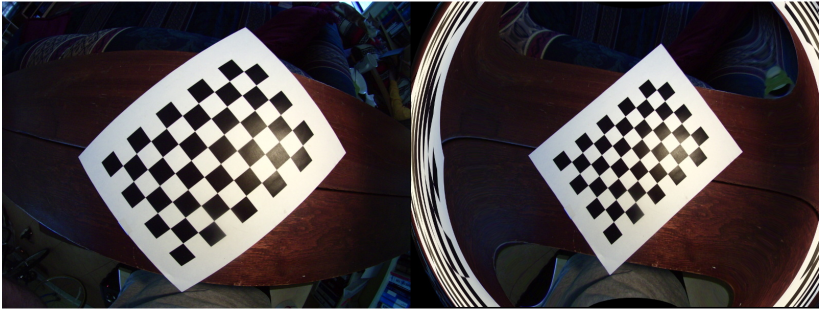
1. we know R_{cb}, d_{cb}, K we can find the projection for some $k = 0, p = 0$
2. minimize the reprojection error

$$k, p = \arg\min_{k,p} \sum_i \left(u^i - \text{proj}\left(v^i, K, k, p\right)\right)^2$$

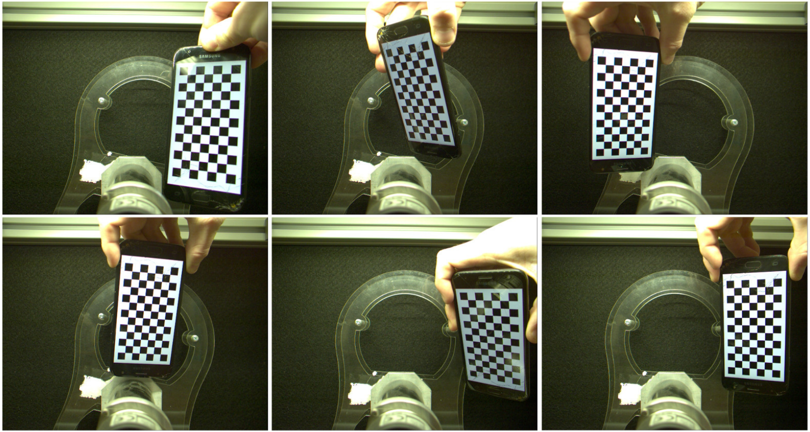
3. undistort points v^i
4. find new K
5. minimize reprojection error
6. etc

Fix Nonlinear Effects

If we know K, k, p, then any image can be transformed as it was captured by a perfect camera



Intrinsics Calibration Steps



Perspective-n-Point Problem

We found intrinsics K, k, p

We are able to reconstruct the rays!

How to find position of an object in camera CS frame?

Position of camera in world CS frame?

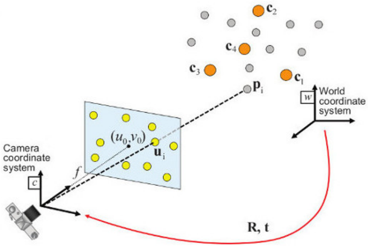
Estimate Object Pose

Perspective-n-Point Problem

Let we know

- world coordinates of points: $p^i = (p^i_x, p^i_y, p^i_z, 1)$
- their projections: $u^i = (u^i_x, u^i_y, 1)$

The goal is to find rotation and displacement: R, r.



system of $2n$ linear equations wrt (R, r) matrix

$$u^i \sim K \begin{pmatrix} R & r \end{pmatrix} p^i$$

$\Rightarrow$ if $(R, r) \in \mathbb{R}^{3\times4}$ we need 6 pairs of points

Perspective-n-Point Problem

On the other hand $(R, r) \in SO(3) \times \mathbb{R}^3 \subset \mathbb{R}^{3\times 4}$

We have 6 more additional constraints: $R^T R = I, \det R = 1$

$\Rightarrow$ 6 indep parameters

$\Rightarrow$ 3 pairs of points are needed

We are given 3 pairs:

1. construct a plane crossing the points p^1, p^2, p^3:

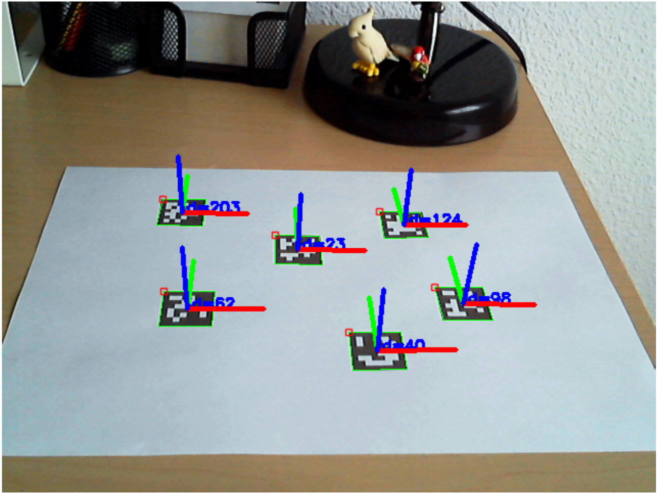
$$(x, y, z)^T = (p_2 - p_1, p_3 - p_1, p_1) \cdot (v_1, v_2, 1)^T = P \cdot v$$

2. find homography $H : v^i \mapsto u^i$
3. $H \sim K \cdot (R, r) \cdot \begin{pmatrix} p_2 - p_1 & p_3 - p_1 & p_1 \\ 0 & 0 & 1 \end{pmatrix}$

Minimize the reprojection error

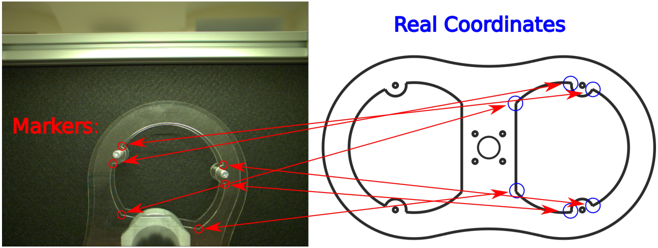
$$R, r = \arg \min_{R \in SO(3), r \in \mathbb{R}^3} \sum_i \operatorname{dist}\left(u^i, K \begin{pmatrix} R & r \end{pmatrix} p^i\right)$$

Example: Aruco Markers Detector



Conclusion: if you know shape of the object, you can find it's pose (if not a degenerate case)

Butterfly Extrinsic Calibration



How to find pose of the camera relative to the plates?

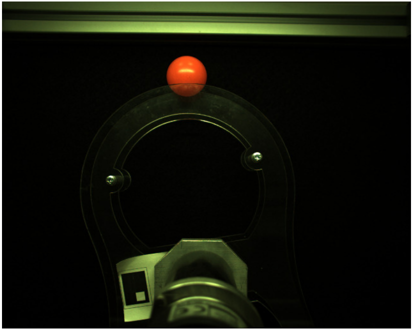
1. points correspondence $p^i \mapsto u^i$
2. solve PnP problem

Estimate Position of a Ball

Estimate Position of a Ball

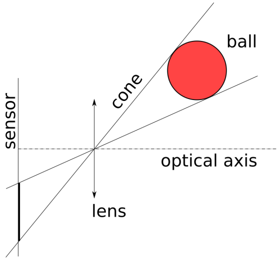
Problem:

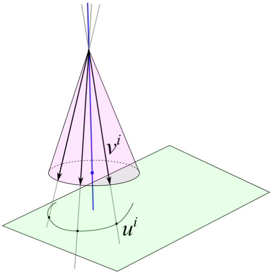
- given a photo
- camera intrinsics
- size of the ball (meters)
- need x, y, z



What is a Projection of a Sphere?

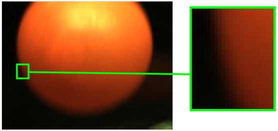
- The rays form a cone
- The cone crosses the plane of sensor
- The section of cone by a plane is an ellipse

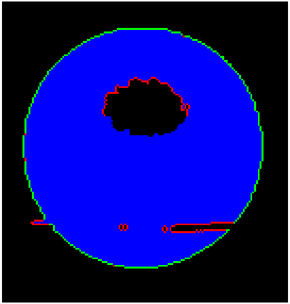




Step 1: Edge Detection

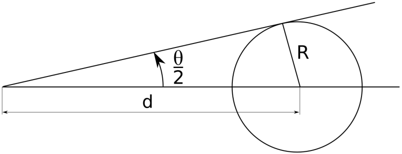
- The rays form a cone
- Edge point to straight line
  $u^i \mapsto a^i$
- Straight line to a unit vector
  $a^i \mapsto v^i$
- All the unit vectors must point to the same plane!





Step 2: Fine Cone Equation

- The rays form a cone
- The main axis of the cone crosses the ball center
- The apex of the cone gives the distance



1. find the plane n fitting by the "points" v^i: $n \cdot v^i = 0$
2. find the cone apex angle: $\theta = 2\frac{\sum_i^N \arccos \frac{n \cdot v^i}{\|n\|}}{N}$
3. find the distance: $\sin \frac{\theta}{2} = \frac{R}{d}$
4. the ball coordinates: $p = \frac{R}{\|n\| \sin \frac{\theta}{2}} n$

Estimate Ball Rotation

- Markers (feature points) on the ball
- Motion vectors

$$u_1 \sim K \cdot p$$
$$u_2 \sim K \cdot T \cdot p$$

then

$$K^{-1} \cdot u_1 \sim p$$
$$T^{-1} \cdot K^{-1} \cdot u_2 \sim p$$

or

$$u_1 \sim \underbrace{K \cdot T^{-1} \cdot K^{-1}}_{H} u_2$$