* Pintura de Rajib Roy, Case Investigation - 2012

# Digital Forensics

*MO447 / MC919*


***Prof. Dr. Anderson Rocha***

*Microsoft Research Faculty Fellow*
*Affiliate Member, Brazilian Academy of Sciences*
Reasoning for Complex Data (Recod) Lab.

anderson.rocha@ic.unicamp.br
http://www.ic.unicamp.br/~rocha

**Reasoning for Complex Data (RECOD) Lab.**
Institute of Computing,
University of Campinas (Unicamp)

Av. Albert Einstein, 1251 - Cidade Universitária
CEP 13083-970 • Campinas/SP - Brasil

# Eye specular highlights telltales for digital forensics

Priscila Saboia, Tiago Carvalho, Anderson Rocha: Eye specular highlights telltales for digital forensics: A machine learning approach. ICIP 2011: 1937-1940

# Is this image fake or real?

# Introduction

- Forensic analysis of digital documents can be used to attest the authenticity of photographs.

- There are many techniques that explore features such as:
  - Compression artifacts;
  - Statistical descriptors;
  - Acquisition telltales;
  - Illumination inconsistencies.

# Introduction

- **Composites of People**: people put together artificially in a single photograph.

- Each person was photographed in different illumination conditions.

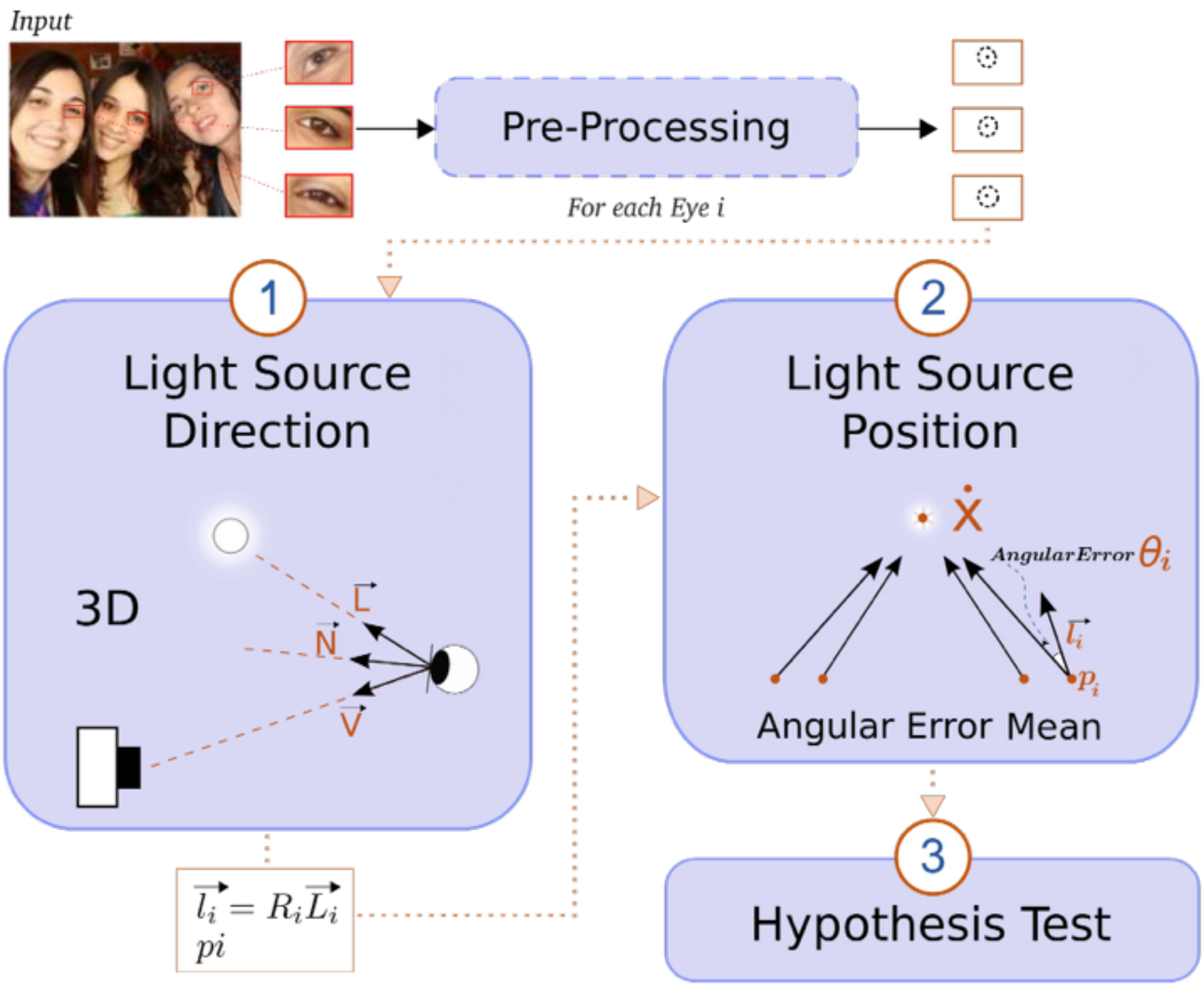
- Illumination inconsistencies can be used to detect composites of people.

# Introduction

- **Illumination inconsistencies**:
  Johnson and Farid (2007) analyzed specular highlights in the eyes of people in a same picture.

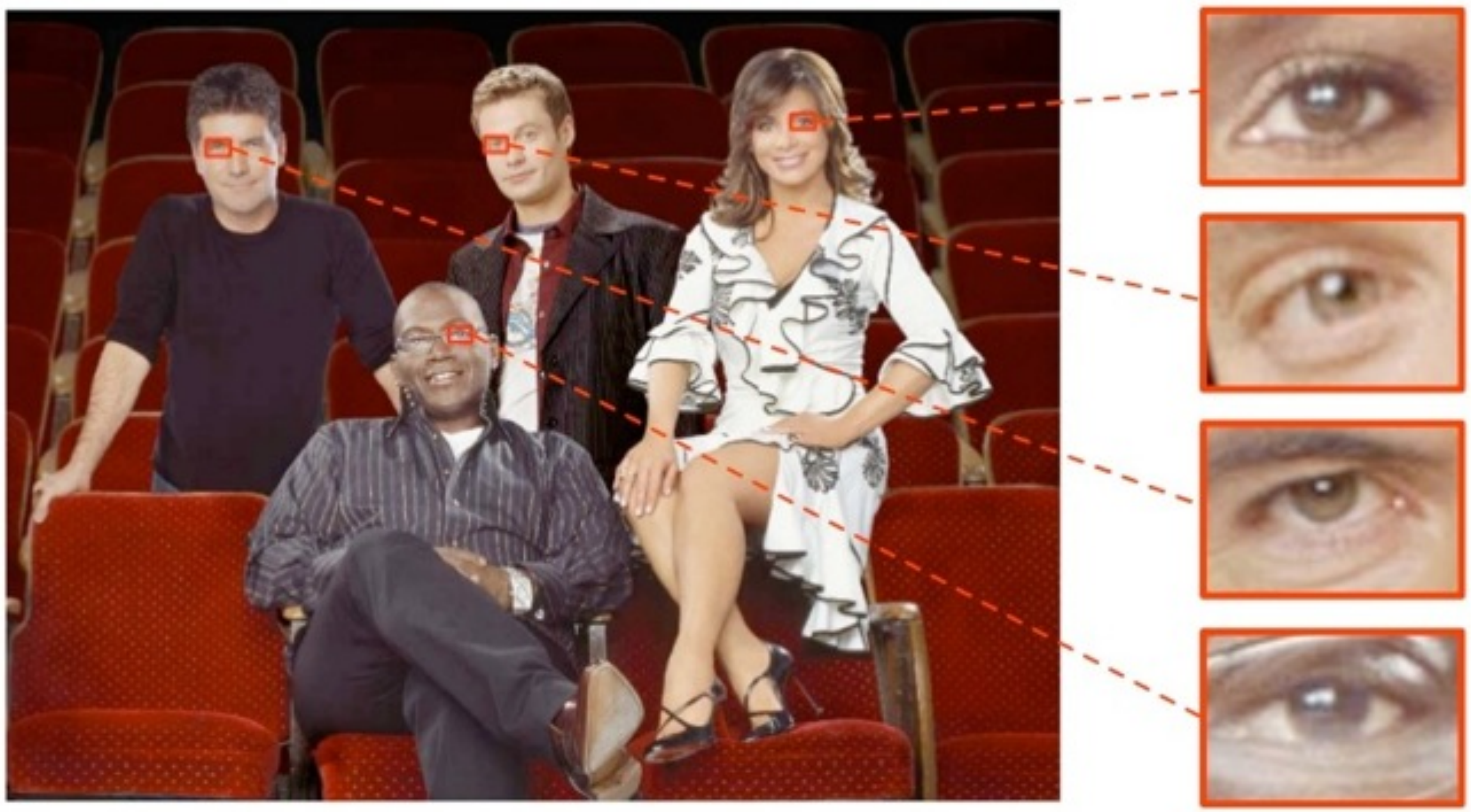
- Our proposal extends upon their work, by the means of:
  - Extraction of new features;
  - Machine learning decision process.

# Johnson and Farid's Method



# Example

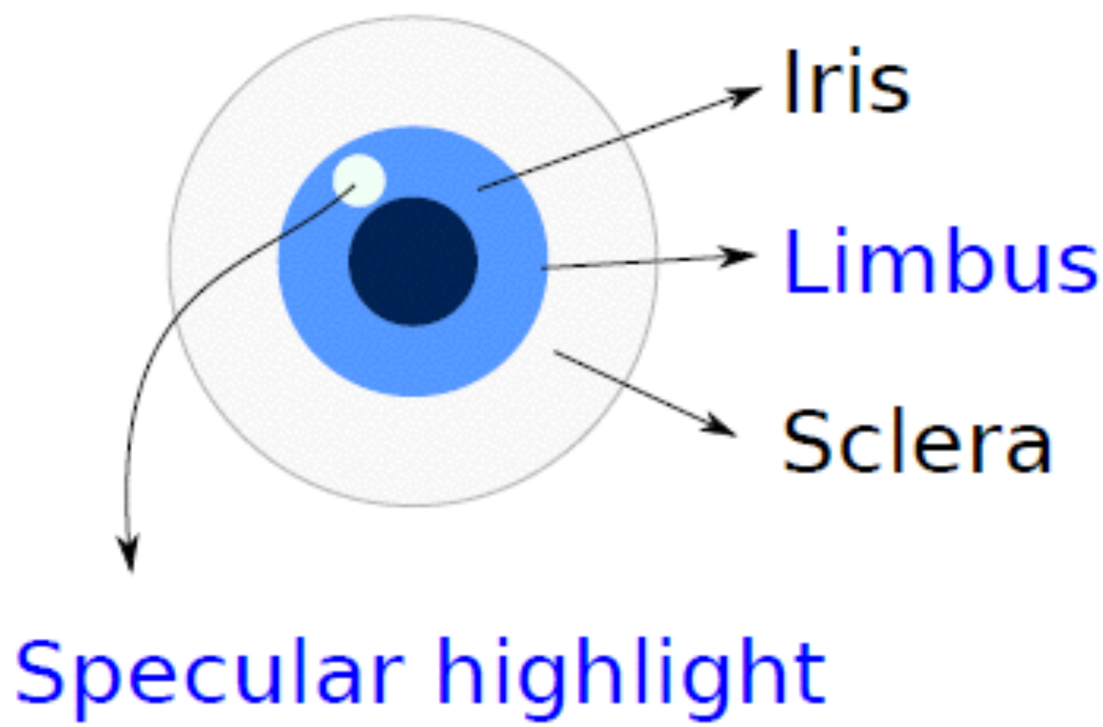
- *American Idol*



© Fox News and the Associated Press

# Pre-processing

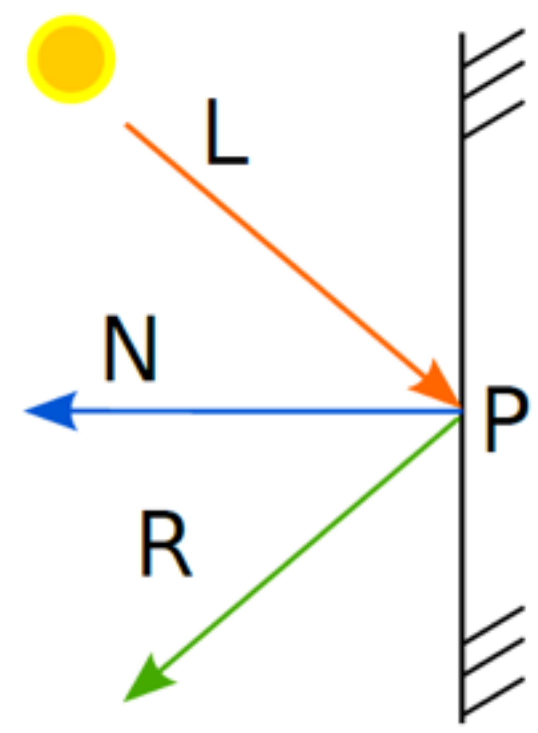
- **Objective**: Extract useful information from a photograph to analyze its authenticity.



# Stage 1 – Light Source Direction

- **Objective:** Estimate the direction connecting the eye to a light source.
- **Fundaments**: Law of reflection

$$L = 2(R^T N)N - R$$



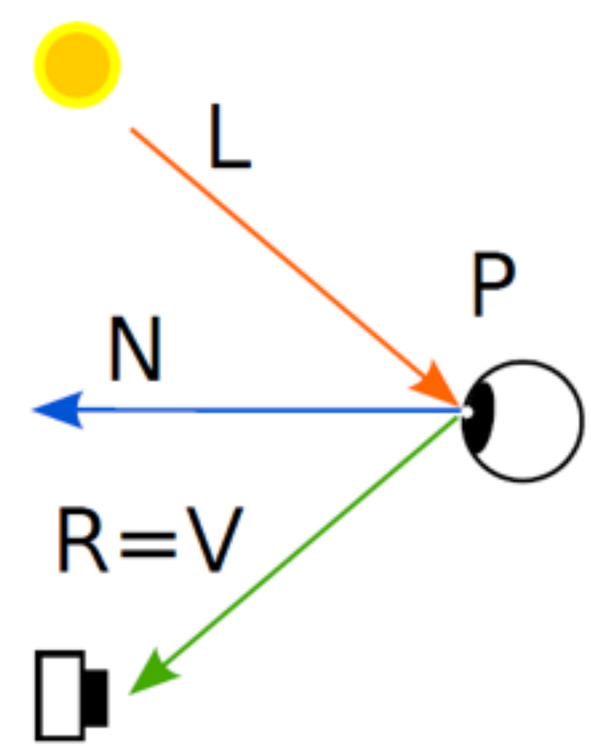
# Stage 1 – Light Source Direction

- **Objective:** Estimate the direction connecting the eye to a light source.
- **Fundaments**: Law of reflection

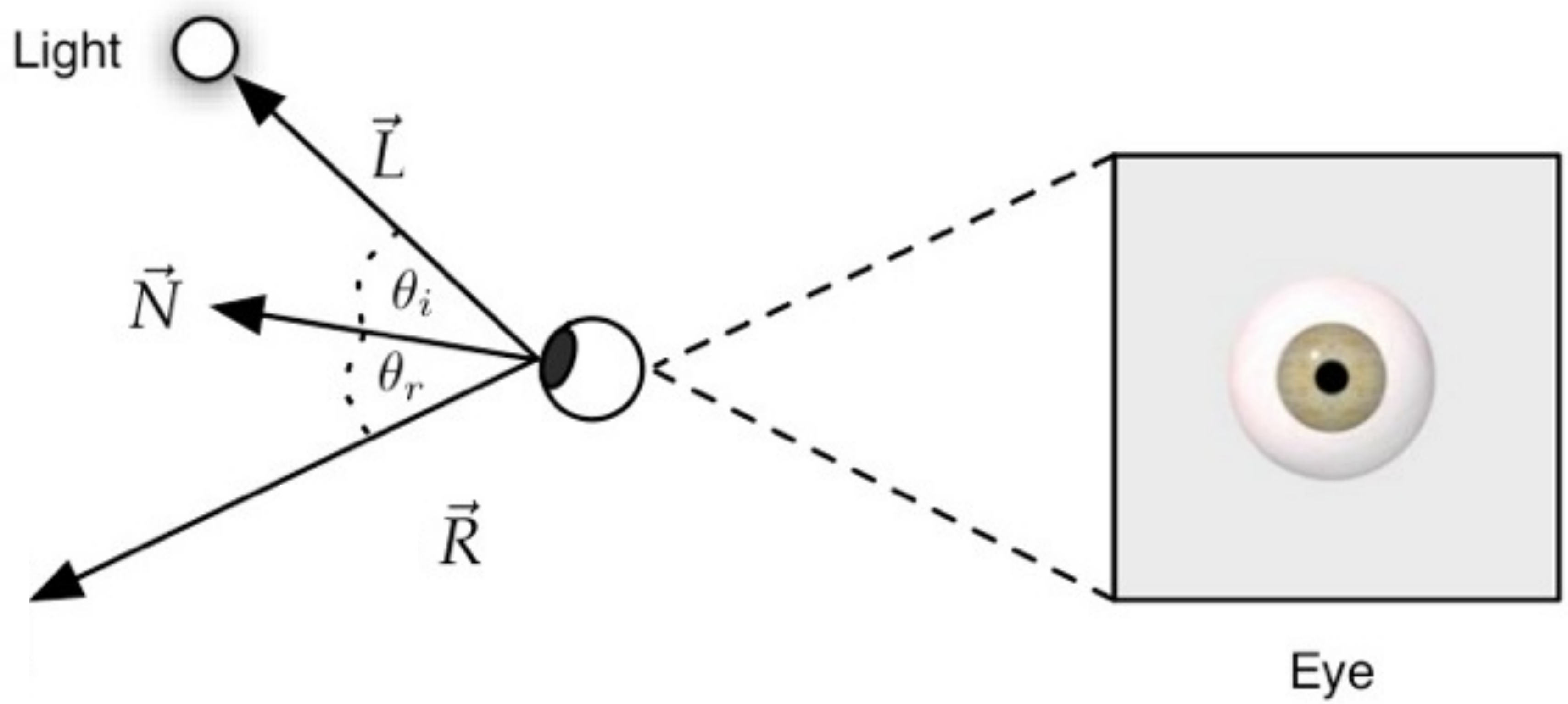
$$L = 2(R^T N)N - R$$

- Assuming the eyes are perfect reflectors, V=R:

$$L = 2(V^T N)N - V$$



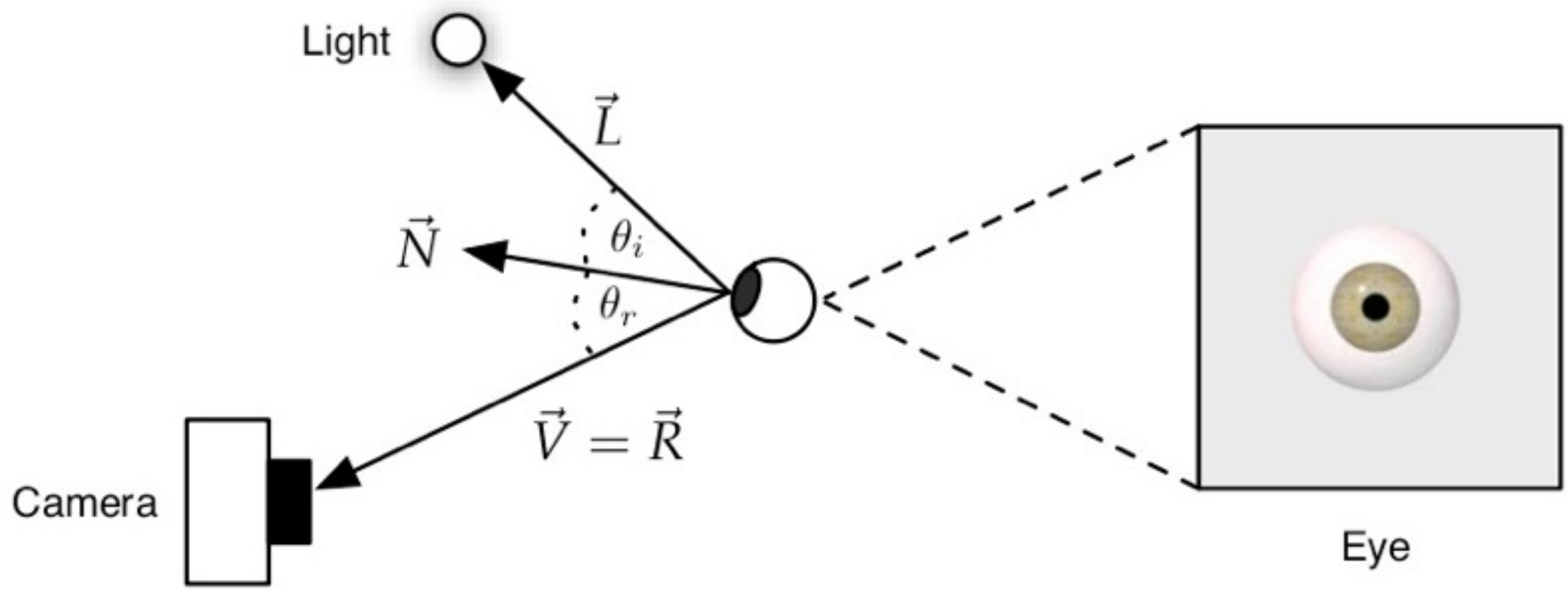
# Stage 1 – Light Source Direction






$$L = 2(V^T N)N - R$$

# Stage 1 – Light Source Direction




© Micah Kimo Johnson

$$L = 2(V^T N)N - V$$

# Stage 2 – Source Light Position

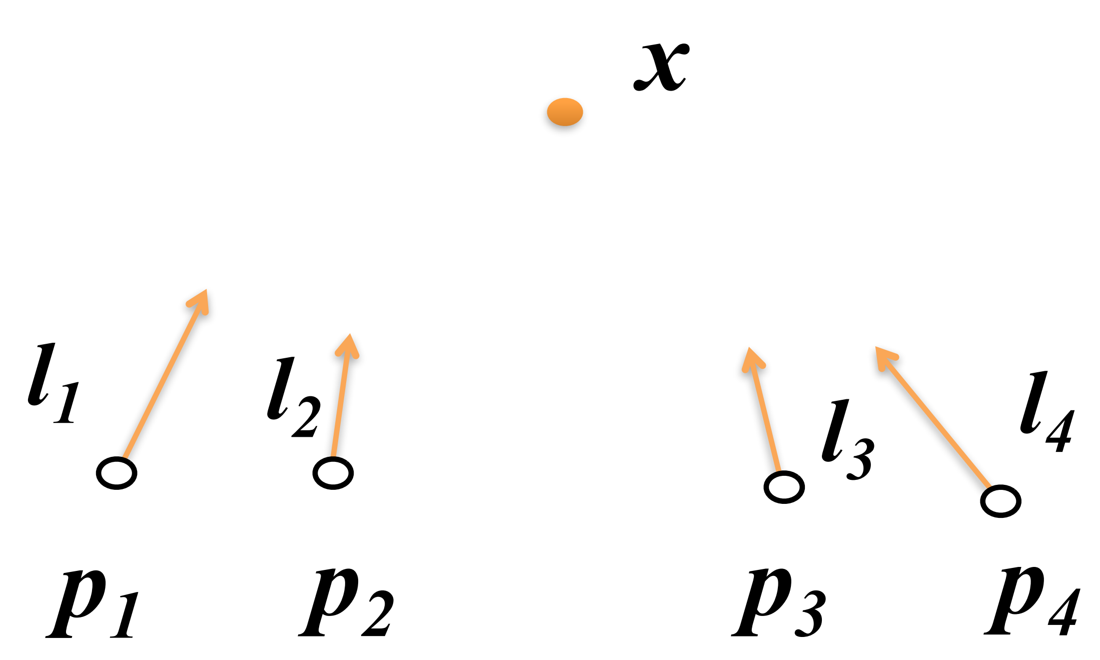
**Objectives:**

- Estimate a single position of the light source for all the eyes.
- Calculate quantities (angular errors) that can be used later to decide whether the image is a composite or not.

# Stage 2 – Source Light Position

**Fundaments:**

- For all the specular highlights $\boldsymbol{p}$, the estimated directions $\boldsymbol{l}$ to their light sources shall converge to a single point $\boldsymbol{x}$.

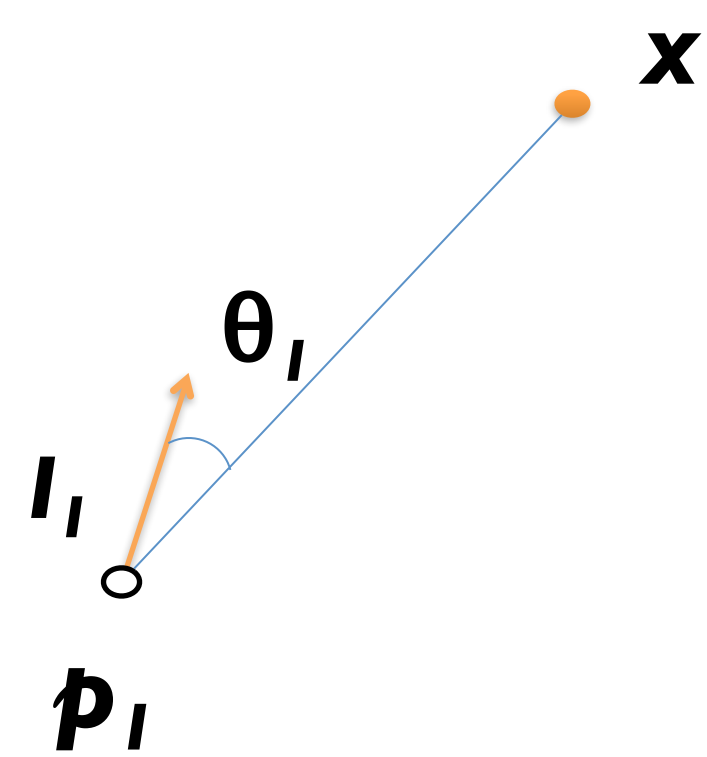


# Stage 2 – Source Light Position

**Fundaments:**

- Given a specific direction ***l**$_l$*, and a freely chosen point ***x*** to serve as a candidate for the light source, one can calculate an **angular error** $\theta_l$ between them:

$$\theta_i(\boldsymbol{x}) = \arccos\left(\boldsymbol{l}_i{}^T \frac{\boldsymbol{x} - \boldsymbol{p}_i}{||\boldsymbol{x} - \boldsymbol{p}_i||}\right)$$



# Stage 2 – Source Light Position

**Fundaments:**

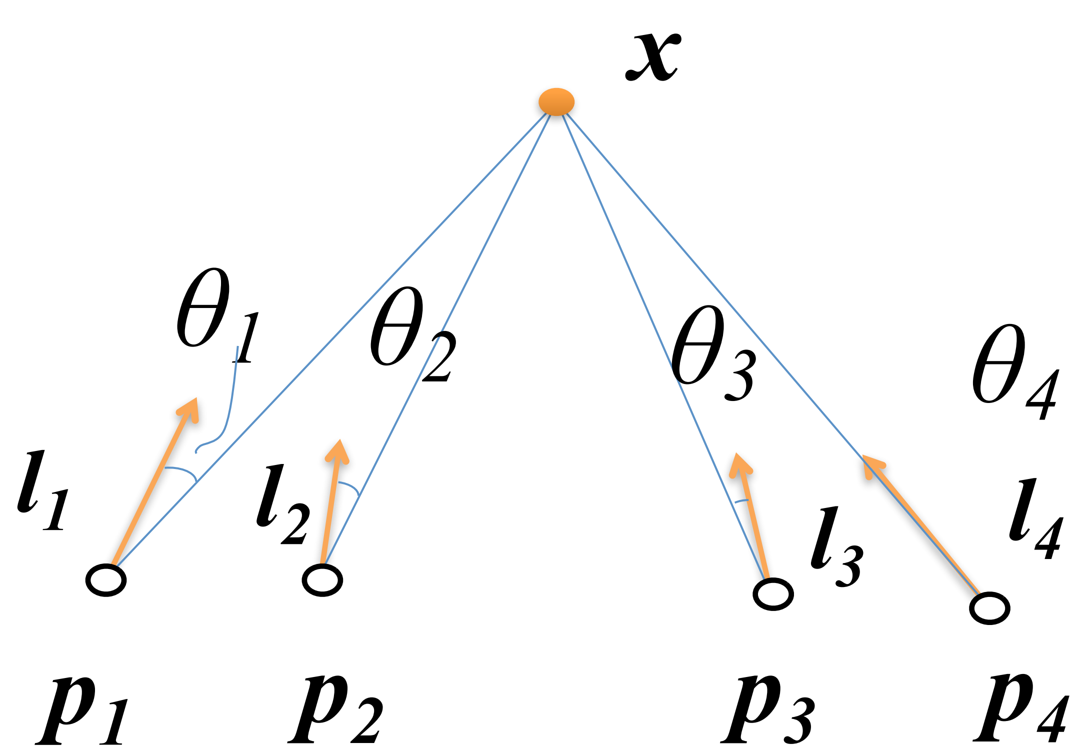
- For an image with ***n*** specular highlights, the best position of the light source can be estimated by minimizing the sum of the angular errors $\theta_i$ .

$$E(\boldsymbol{x}) = \sum_{i=1}^{n} \theta_i(\boldsymbol{x})$$

# Stage 2 – Source Light Position

## Fundaments:

- For an image with ***n*** specular highlights, the best position of the light source can be estimated by minimizing the sum of the angular errors $\theta_I$ .

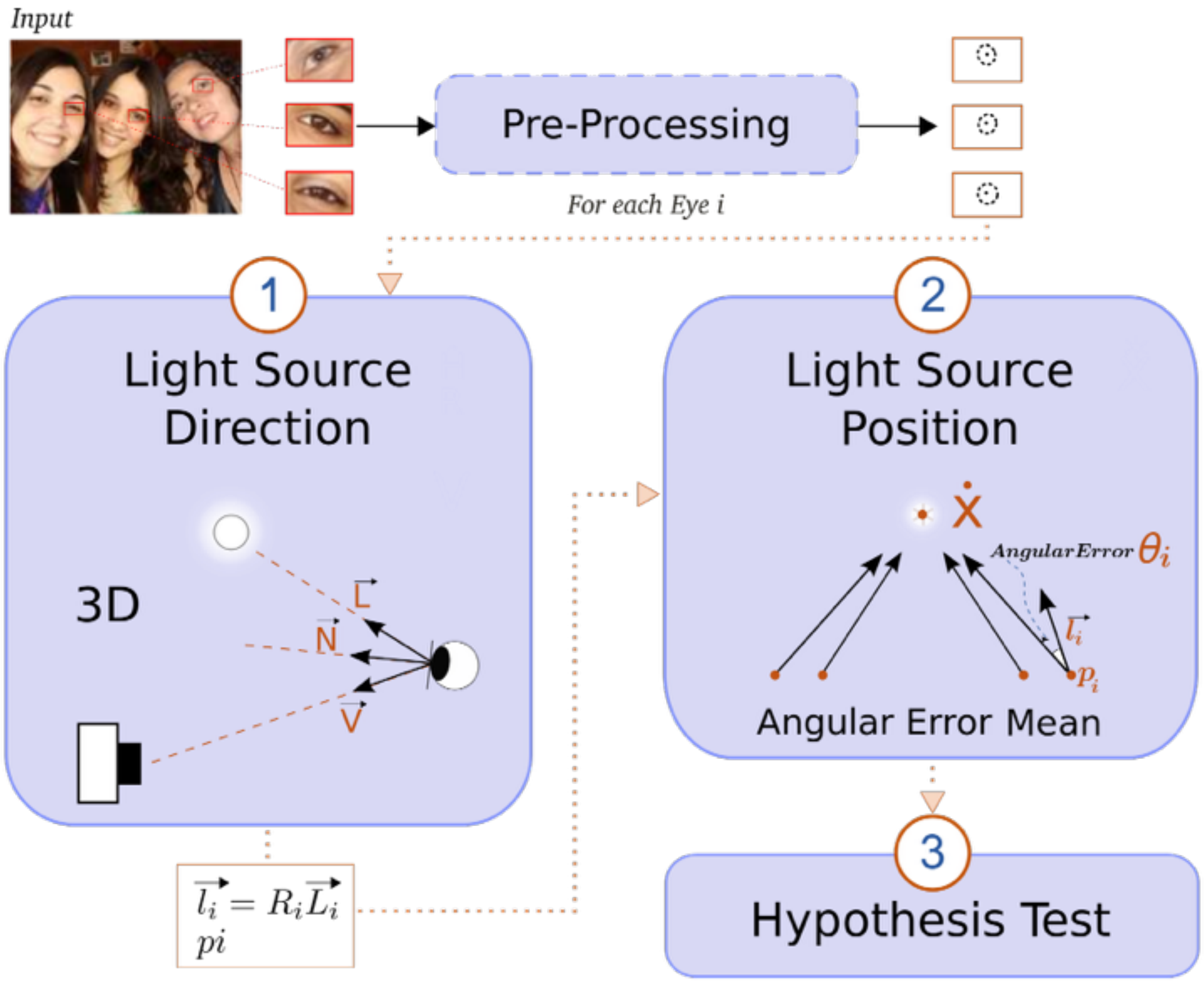


# Stage 3 – Decision

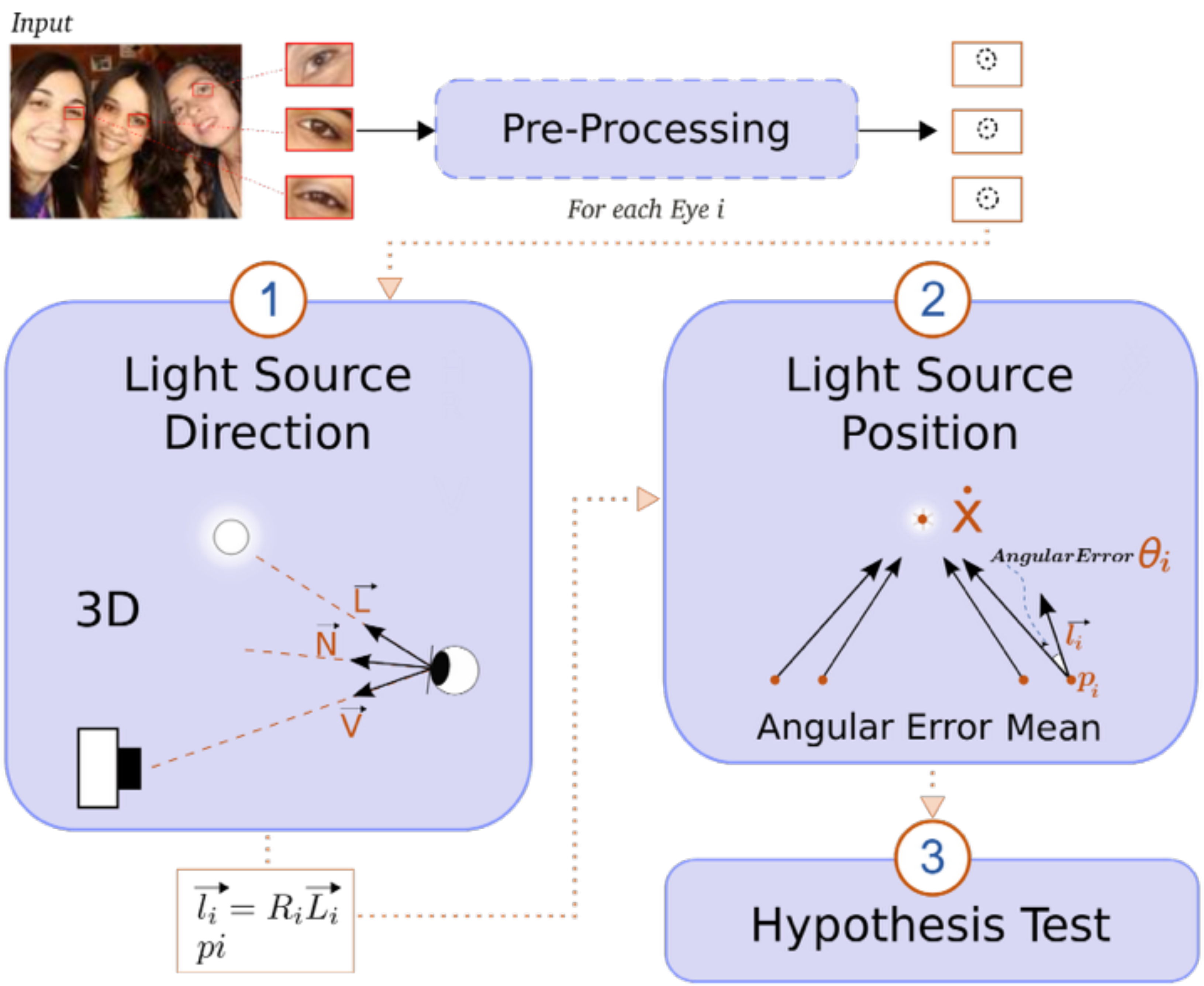
**Objective:**

- Decide whether or not an image is tampered, by the application of classical hypotheses tests over the **mean value of the angular errors**.

# Our proposal



# Our proposal

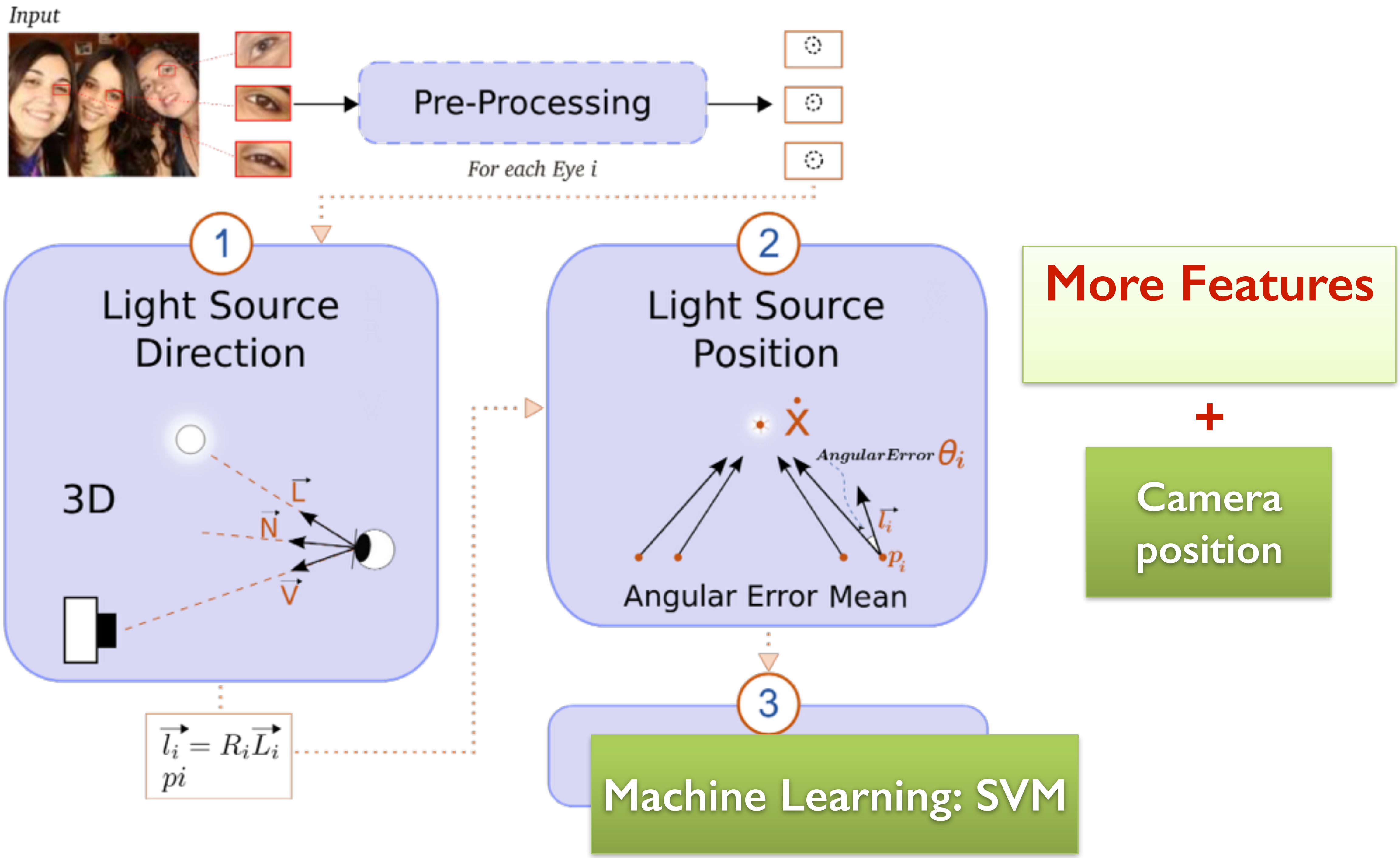



**More Features**

**+**

**Camera position**

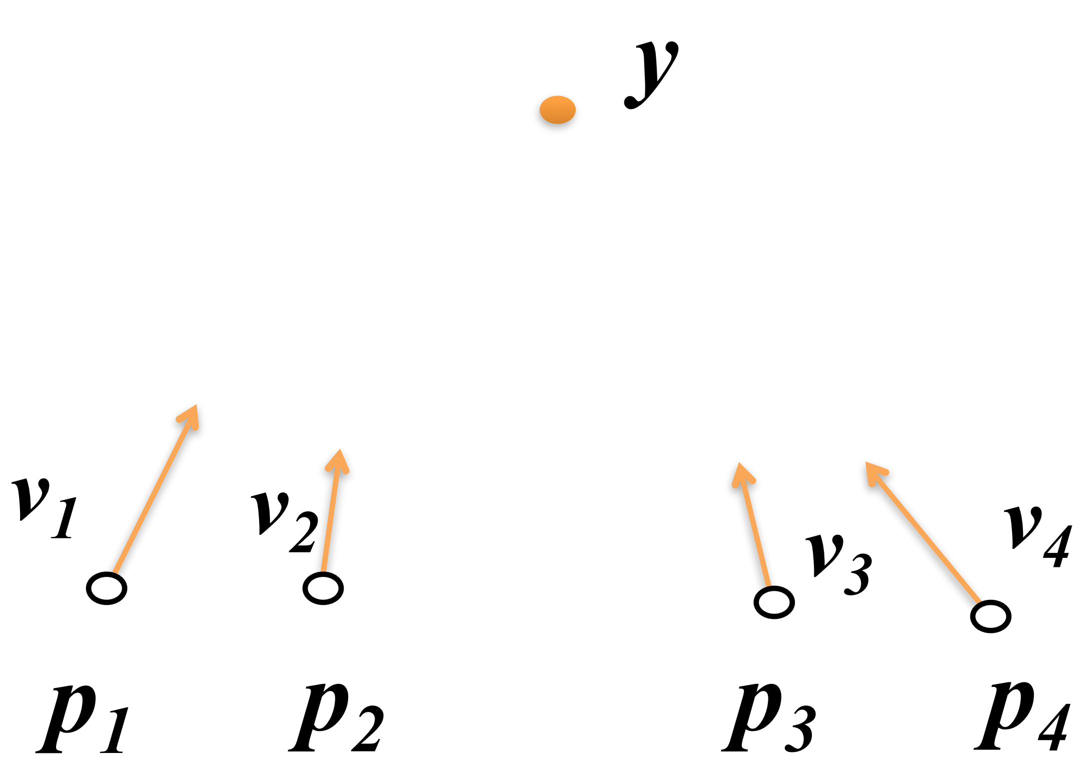
# Our proposal



# Stage 2 – Camera Position

**Fundaments:**

- For all the specular highlights $\boldsymbol{p_i}$ , the estimated directions $\boldsymbol{v_i}$ to their capture points shall converge to a single point $\boldsymbol{y}$.

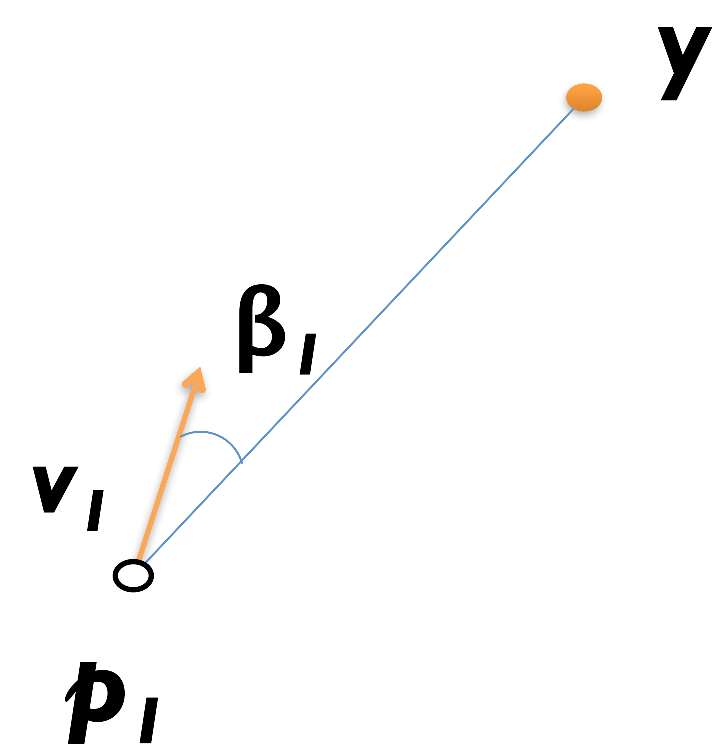


# Stage 2 – Camera Position

## Fundaments:

- Given a specific direction $\boldsymbol{v_i}$, and a freely chosen point **y** to serve as a candidate for the position of the camera, one can calculate an angular error $\beta_i$ between them:

$$\beta_i(\boldsymbol{y}) = \arccos\left(\boldsymbol{v}_i{}^T \frac{\boldsymbol{y} - \boldsymbol{p}_i}{||\boldsymbol{y} - \boldsymbol{p}_i||}\right)$$



# Stage 2 – Camera Position

**Fundaments:**

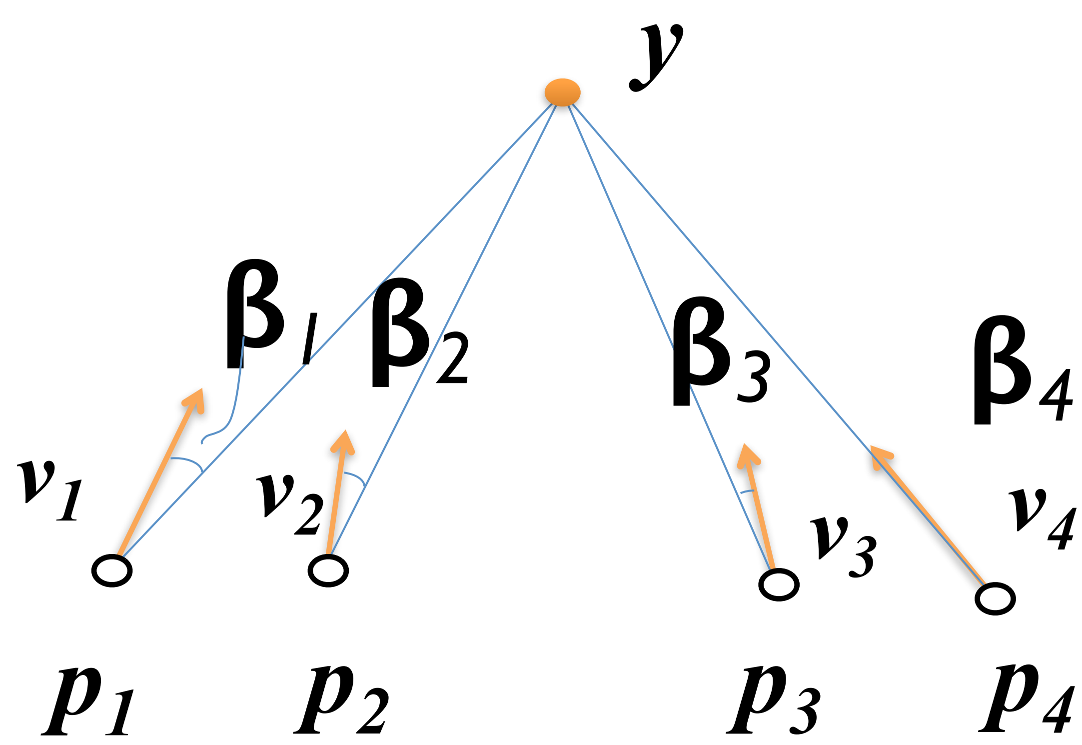
- For an image with ***n*** specular highlights, the best position of the camera can be estimated by minimizing the sum of the angular errors $\boldsymbol{\beta}_i$.

$$E(\boldsymbol{y}) = \sum_{i=1}^{n} \beta_i(\boldsymbol{y})$$

# Stage 2 – Camera Position

## Fundaments:

- For an image with ***n*** specular highlights, the best position of the camera can be estimated by minimizing the sum of the angular errors $\beta_i$.



# Stage 2 – Camera Position

## Stage 3:

- Use four features to decide about authenticity images:

**LME** - The mean value of the angular errors related to the light source position (maintained from Johnson and Farid's Method);

**LSE** - The standard deviation of the angular errors related to the light source position;

**VME** - The mean value of the angular errors related to the camera position;

**VSE** - The standard deviation of the angular errors related to the camera position.

# Stage 3 – Our Proposal

**Stage 3:**

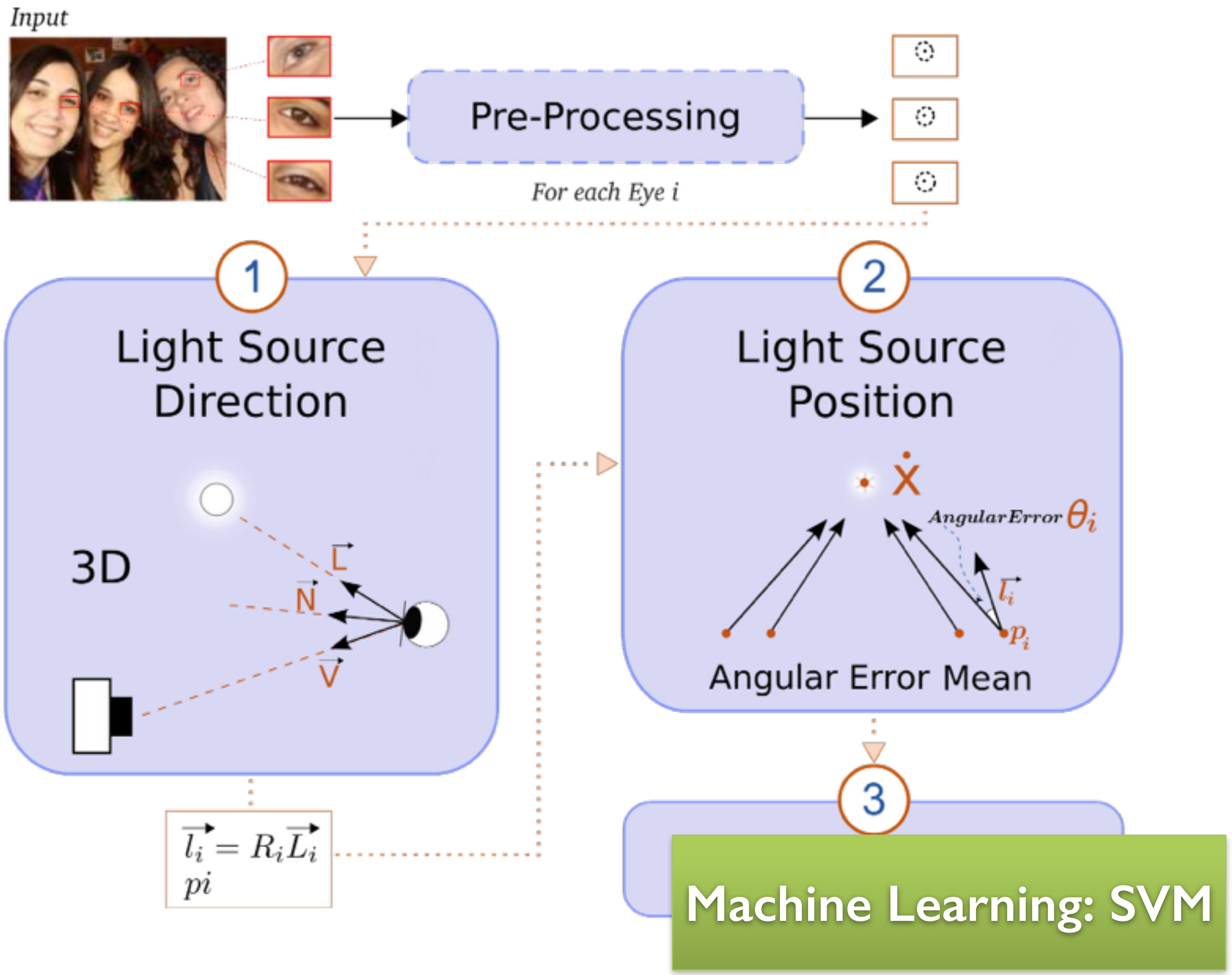
- Decide whether or not the image is a composite, based on the four LME, LSE, VME and VSE features;
- Apply SVM-based machine learning techniques instead of a hypotheses test, in such process.

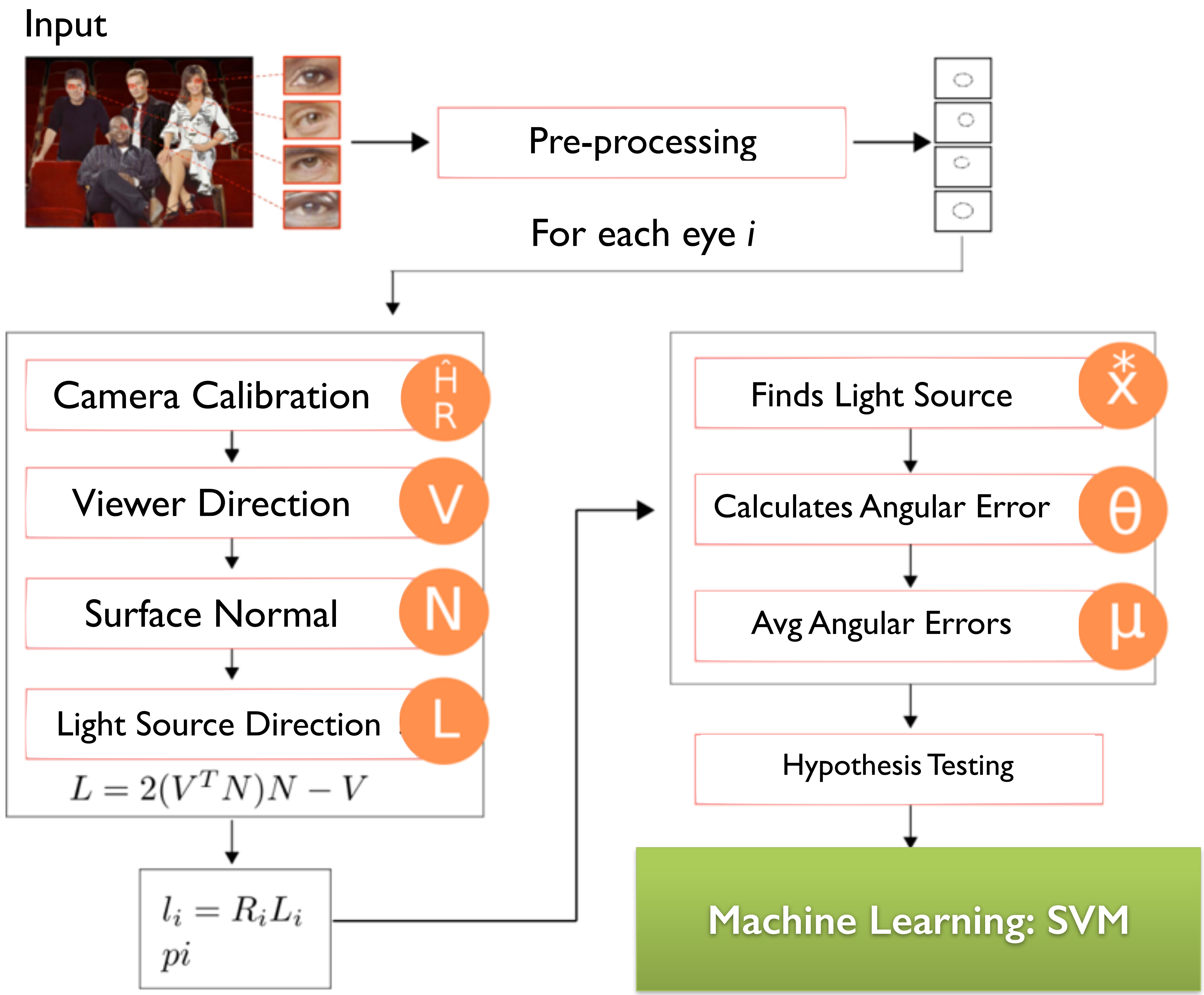
# The Devil is in the Details

# Coordinate Systems

- We will be dealing with two different coordinate systems
  - System relative to the position of the eye
    - Uppercase
  - System relative to the position of the camera
    - Lowecase

# Our proposal




Input
Pre-processing
For each eye *i*
Camera Calibration
Ĥ
R
Viewer Direction
V
Surface Normal
N
Light Source Direction
L
$L = 2(V^T N)N - V$
$l_i = R_i L_i$
$pi$
Finds Light Source
X̊
Calculates Angular Error
θ
Avg Angular Errors
μ
Hypothesis Testing
Machine Learning: SVM

# Step 1:
# Light Source Direction

# Step 1: Light Direction

## Camera Calibration

- For obtaining the source light **L**, it is necessary to know the vectors **V** (viewer's direction) and **N** (surface normal)
- For obtaining **V** and **N** in a common coordinate system, first it is necessary to estimate the projective matrix transformation **H** which describes the transformation from world coordinates to image coordinates

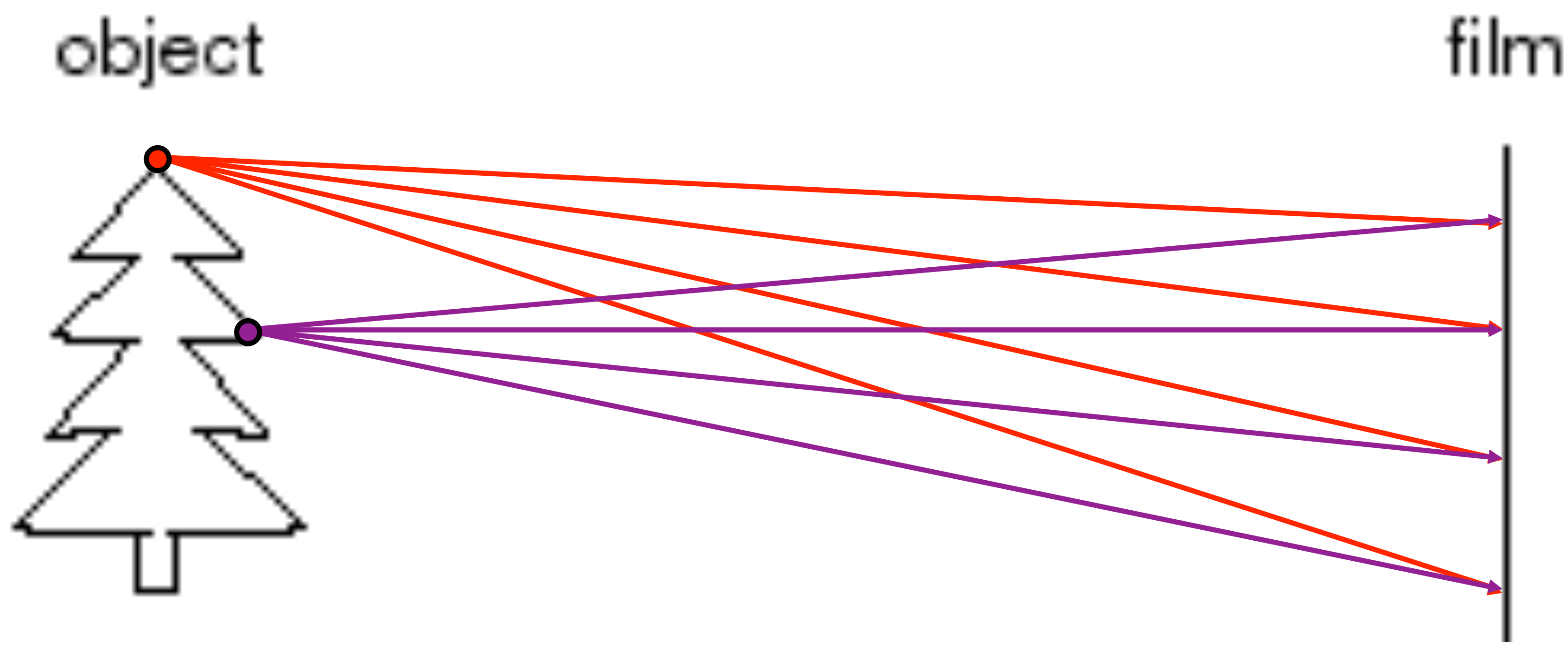
# Step 1: Light Direction
## Camera Calibration

- The *projective transformation* **H** which describes the coordinate system transformation is given by

$$x_{im} = HX$$

$$H = \begin{bmatrix} fs_x & f\tau & c_x \\ 0 & fs_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R \;\; T]$$

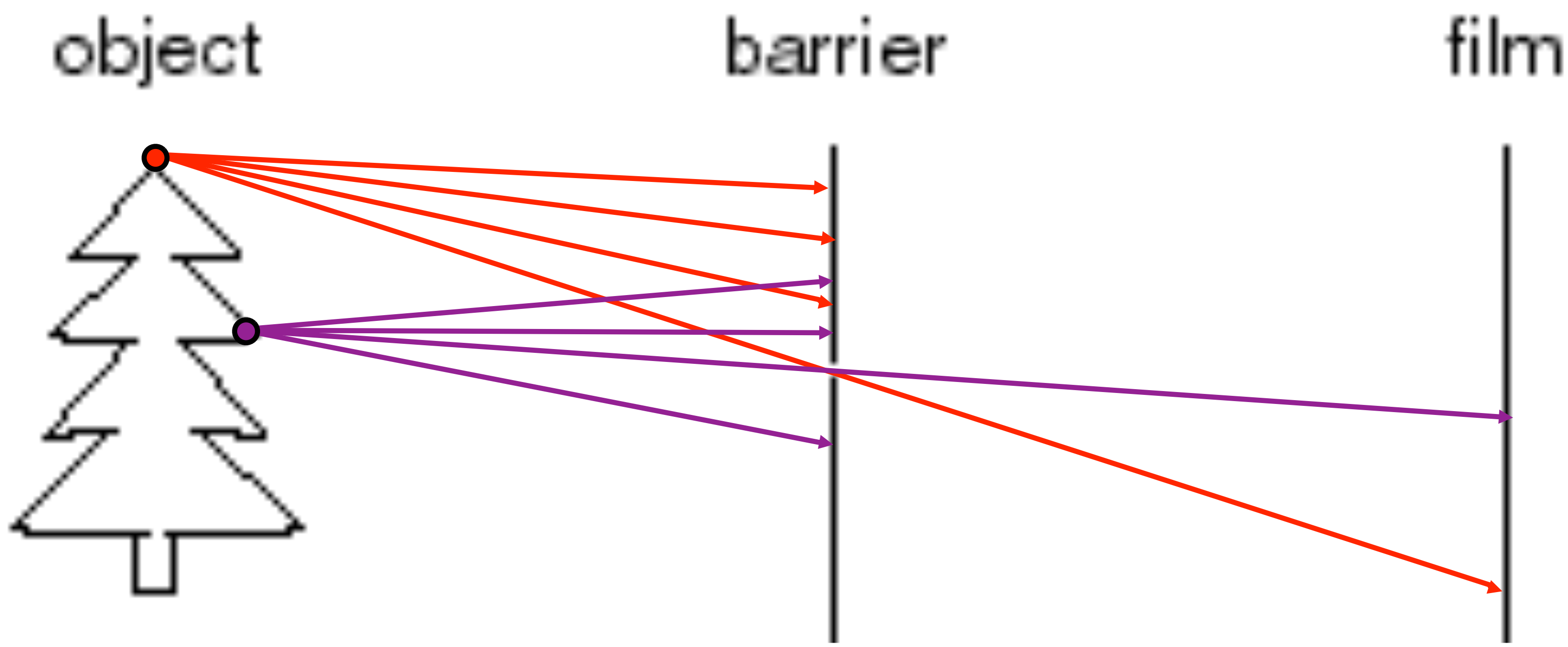
# From Computer Vision 101 :-)




Let's design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

# From Computer Vision 101 :-)

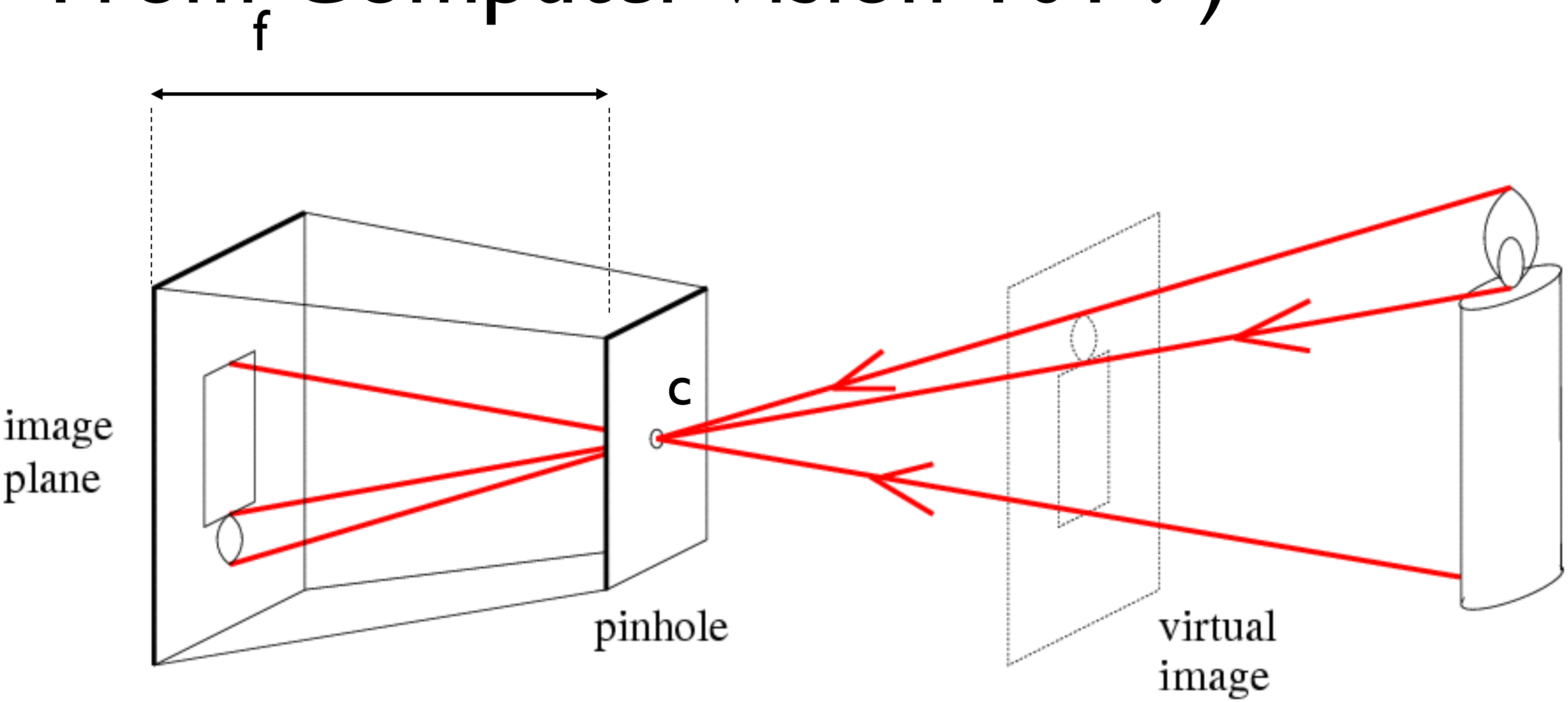



Idea 2: add a barrier to block off most of the rays

– This reduces blurring

– The opening known as the **aperture**

# From Computer Vision 101 :-)

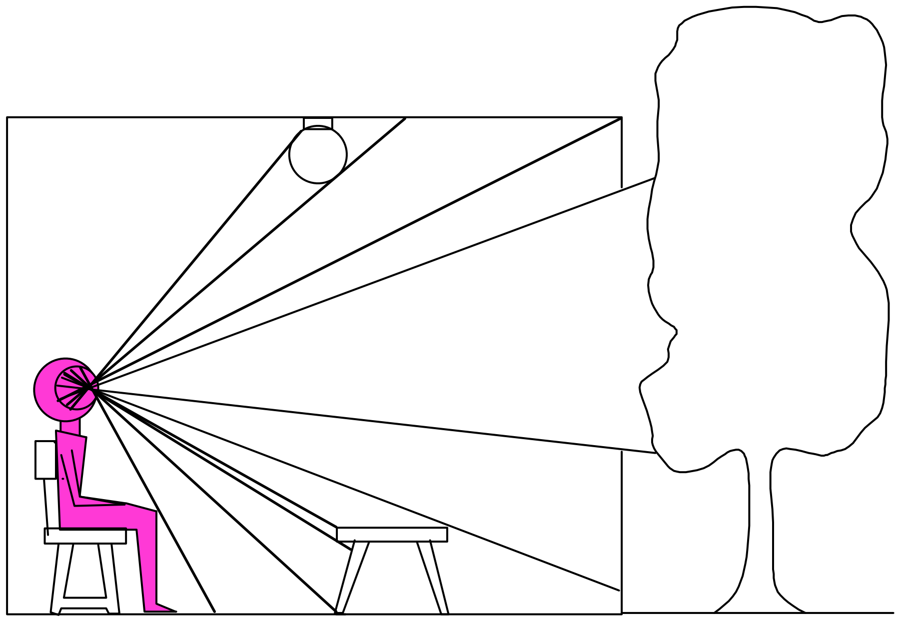


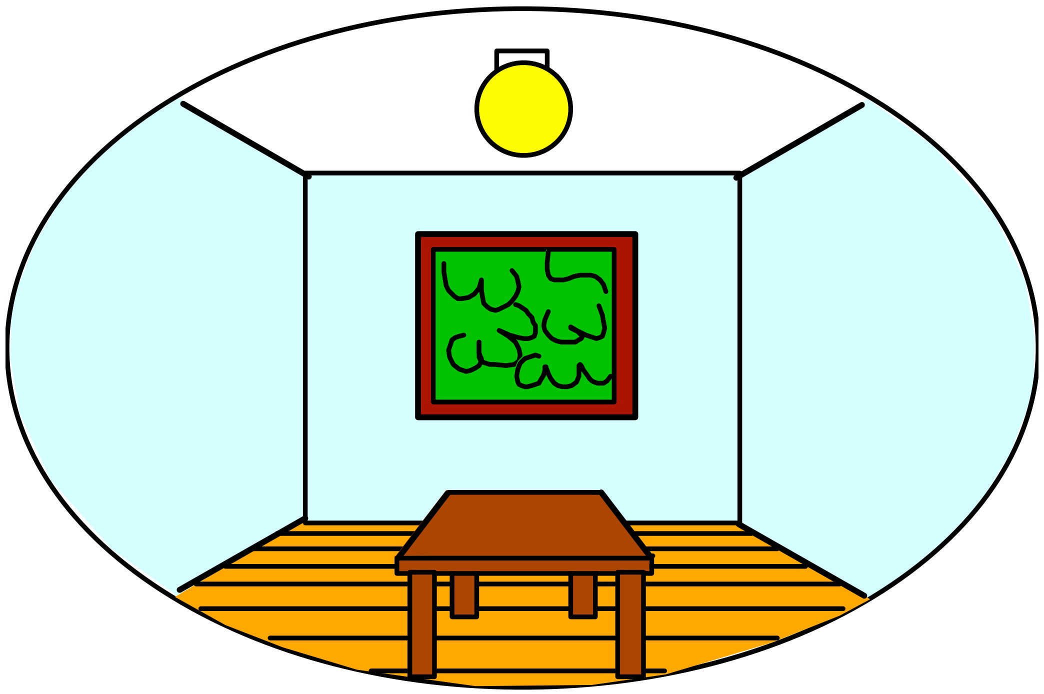

f = focal length
c = center of the camera

# Dimensionality Reduction Machine (3D to 2D)

*3D world*

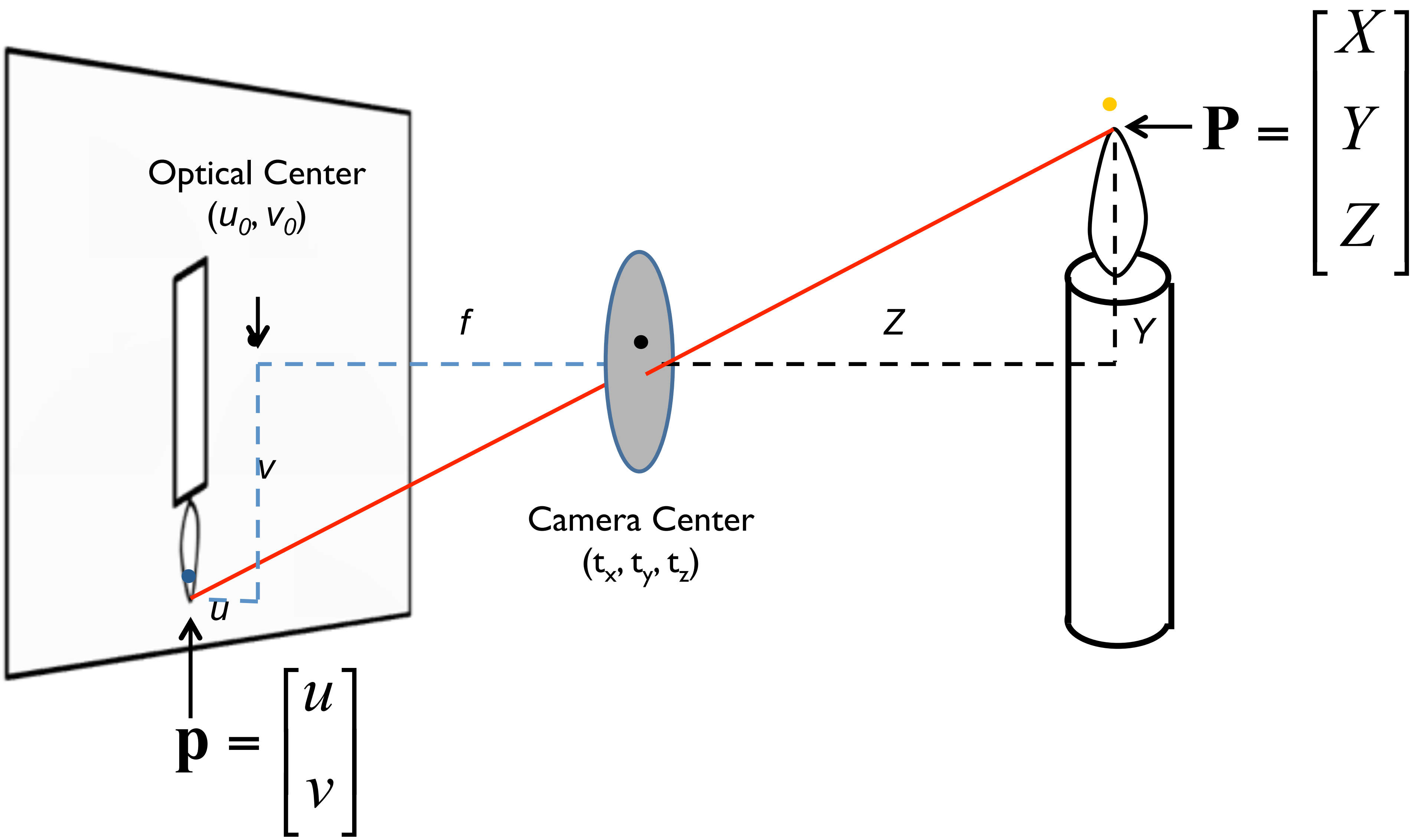
*2D image*



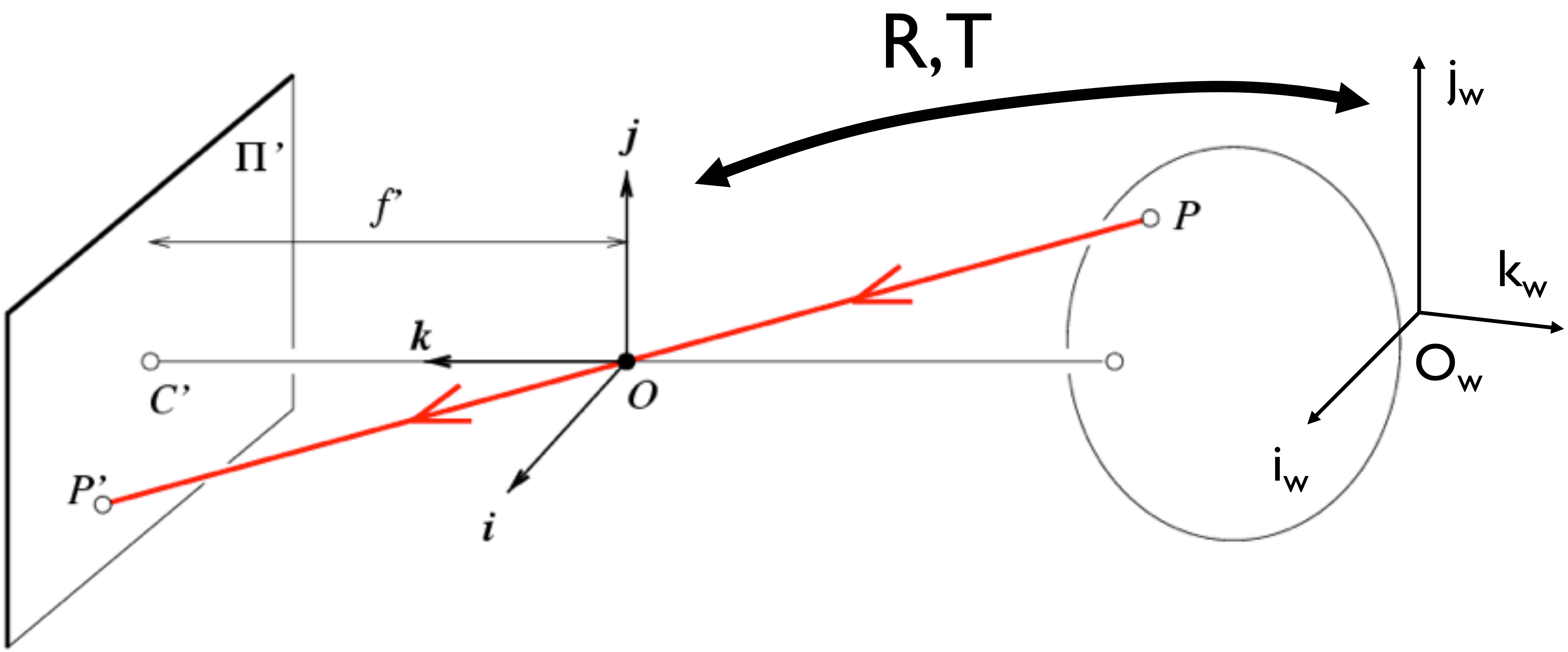


Point of observation

# Projection: world coord. → image coord.



# Projection Matrix



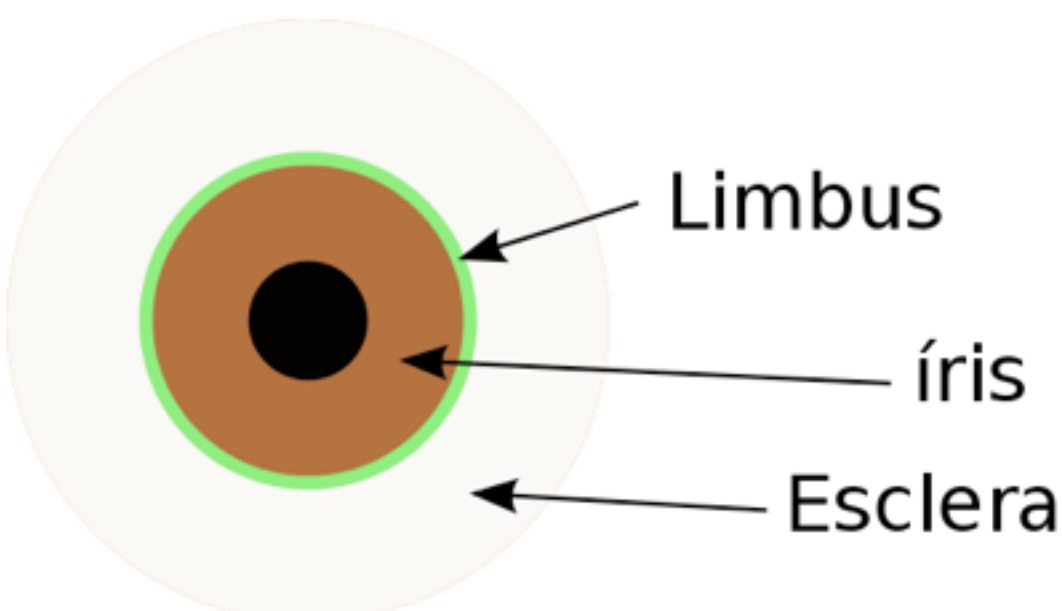

$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

**x**: Image Coordinates: (u,v,1)
**K**: Intrinsic Matrix (3x3)
**R**: Rotation (3x3)
**t**: Translation (3x1)
**X**: World Coordinates: (X,Y,Z,1)

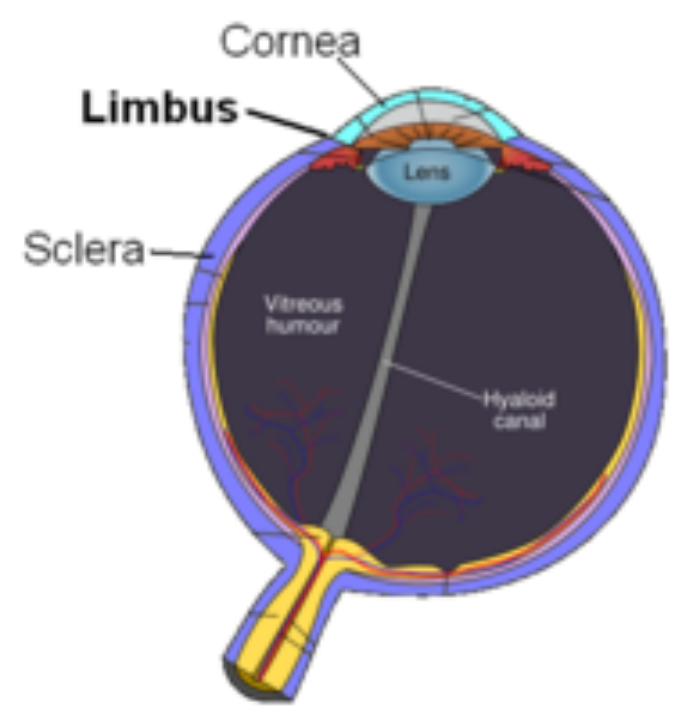
# Step 1: Light Direction

## Camera Calibration

- The eye geometry is explored to estimate the transformation
- The **limbus** is the border between sclera (white part) and the iris (colored part). It can be modeled as a circle (in green in the figure)



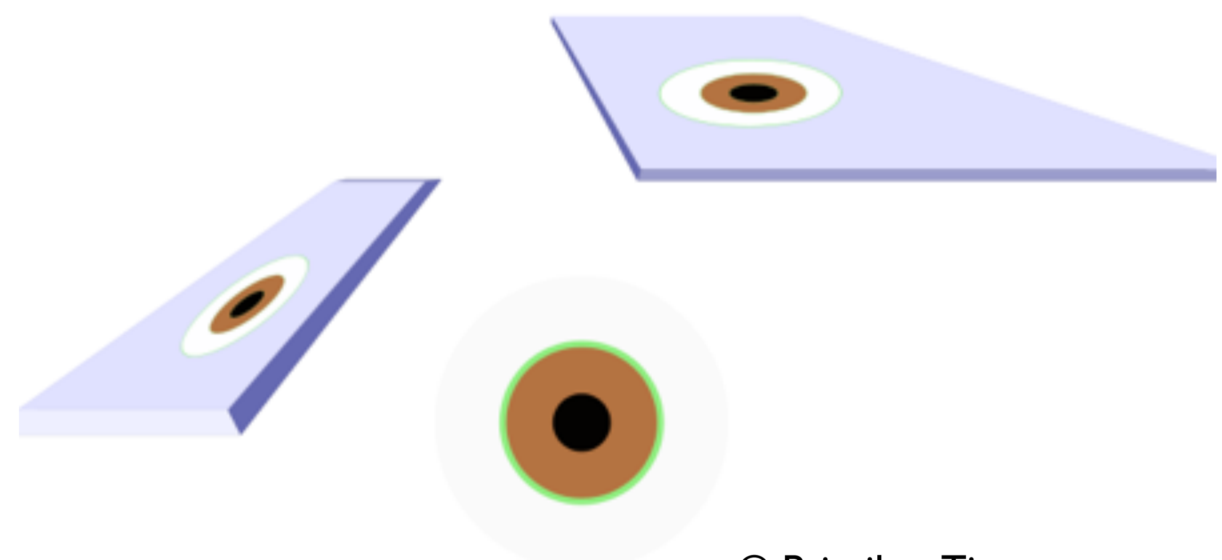

© Priscila and Tiago




© Rhcastilhos

# Step 1: Light Direction
## Camera Calibration

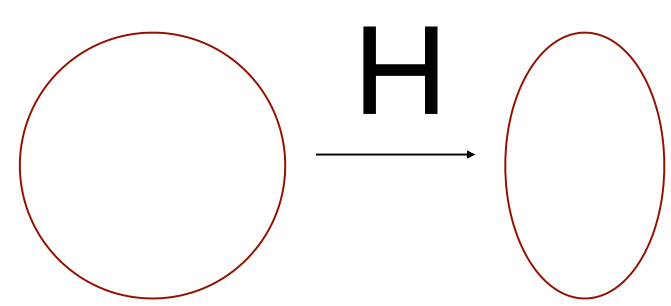
- However, the image of the limbus will be an ellipsis when captured by the camera. The exception is when the eyes are exactly in front of the camera



# Step 1: Light Direction

## Camera Calibration

- Intuitively, the distortion of an ellipsis for a circle is related with pose and position of the eyes with respect to the camera
- In this way, we need a transformation that aligns the image of the limbus to a circle




$$x_{im} = HX$$

# Step 1: Light Direction
## Camera Calibration

- In general, the projective transformation (3D world coord to 2D image coord) are represented by homogeneous coordinates, in a 3x4 matrix H
- We assume the points in the limbus are co-planar and define the world coordinates in such a way they have Z = 0

# Step 1: Light Direction
## Camera Calibration

- With this assumption, the projective matrix H is reduced to a simple 3x3 planar projection transformation where world points *X* and image points *x* are represented by 2D homogenous vector
- Therefore, points in the limbus in world coordinates must obey the following implicit equation of a circle:

$$f(X, \alpha) = (X_1 - C_1)^2 + (X_2 - C_2)^2 - r^2 = 0$$

# Step 1: Light Direction

## Camera Calibration

- Therefore, points in the limbus in world coordinates must obey the following implicit equation of a circle:

$$f(X, \alpha) = (X_1 - C_1)^2 + (X_2 - C_2)^2 - r^2 = 0$$

- The vector $\alpha$ is composed by the center of the circle C and its radius r

$$\alpha = (C_1 \; C_2 \; r)^T$$

# Step 1: Light Direction

## Camera Calibration

- H and $\alpha$ are found using the following error function in these two parameters:

$$E(\alpha, H) = \sum_{i=1}^{m} \min_{X^*} \|x_i - HX_i^*\|^2$$

- This is a non-linear error function (non-linear least square problem), which can be solved using an iterative method (e.g., Gauss-Newton, Levenberg-Marquadt)

# Step 1: Light Direction

## Camera Calibration

- **Problem:** with a single circle, there is no unique transformation H which minimizes the equation
- **Solution**: with two coplanar circles, the transformation can be uniquely determined
- In this way, the error function must incorporate the two eyes for estimating H.

# Step 1: Light Direction
## Camera Calibration

- After estimating H, it must be decomposed in terms of the intrinsic and extrinsic camera parameters

Intrinsic

$$H = \begin{bmatrix} fs_x & f\tau & c_x \\ 0 & fs_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R \;\; T]$$

Extrinsic

# Step 1: Light Direction
## Camera Calibration

- The camera intrinsic parameters
  - Focal lengh f
  - Camera center
  - Geometric distortion introduced by the optical system
  - Pixel ration (relation width/length)

# Step 1: Light Direction

## Camera Calibration

- Simplification:
  - The camera center is the center of the image
  - There is no radial distortion
  - Pixel ratio is 1.
- We are left with only the focal length f

# Step 1: Light Direction

## Camera Calibration

- Camera extrinsic parameters:
  - Rotation matrix **R** and translation vector **t**.
- Together they define the transformation between the world coordinate system and the image coordinate system

# Step 1: Light Direction
## Camera Calibration

- Given that the points in the limbus belong to a unique plane (coplanar), H can be decomposed in terms of its intrinsic and extrinsic parameters

$$H = \lambda K \begin{pmatrix} r_1 & r_2 & t \end{pmatrix} \quad K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- where $\lambda$ is a scale factor, the column vector **r1** e **r2** are the first and second column of the rotation matrix R, **t** is the translation vector and **K** is a 3x3 diagonal matrix with the focus

# Step 1: Light Direction

## Camera Calibration

- Recalling: **H** transforms world coordinates to image coordinates
- We want to estimate the transformation $\hat{H}$ which transforms the world coordinates in camera coordinates and the rotation R.

# Step 1: Light Direction

## Camera Calibration

- With the focal length *f* known, it is possible to estimate $\hat{H}$ directly:

$$\frac{1}{\lambda}K^{-1}H = \left( \begin{array}{ccc} r_1 & r_2 & t \end{array} \right)$$

$$\hat{H} = \left( \begin{array}{ccc} r_1 & r_2 & t \end{array} \right) = \frac{1}{\lambda}K^{-1}H$$

$$R = \left( \begin{array}{ccc} r_1 & r_2 & r_1 \times r_2 \end{array} \right)$$

# Step 1: Light Direction

## Camera Calibration

- With an unknown *f*, we need to estimate it before anything
- For that, we can decompose the H matrix in 8 variables:
  - Focal length
  - Scale factor $\lambda$
  - Rotation angles for x, y e z;
  - x, y and z coordinates for translation t

# Step 1: Light Direction
## Camera Calibration

- Estimating *f*

$$H = \lambda \begin{pmatrix} fc_yc_z & fc_ys_z & ft_x \\ f(s_xs_yc_z - c_xs_z) & f(s_xs_ys_z + c_xc_z) & ft_y \\ c_xs_yc_z + s_xs_z & c_xs_ys_z - s_xc_z & t_z \end{pmatrix}$$

where $c_x = \cos(\theta_x)$ e $s_x = \sin(\theta_x)$, $c_y = \cos(\theta_y)$ e $s_y = \sin(\theta_y)$, e $c_z = \cos(\theta_z)$ e $s_z = \sin(\theta_z)$.

$$H_{2x2} = \begin{pmatrix} \lambda fc_yc_z & \lambda fc_ys_z \\ \lambda f(s_xs_yc_z - c_xs_z) & \lambda f(s_xs_ys_z + c_xc_z) \end{pmatrix}$$

# Step 1: Light Direction

## Camera Calibration

$$H_{2x2} = \begin{pmatrix} \lambda f c_y c_z & \lambda f c_y s_z \\ \lambda f(s_x s_y c_z - c_x s_z) & \lambda f(s_x s_y s_z + c_x c_z) \end{pmatrix}$$

$$\hat{f} = \lambda f$$

$$H_{2x2} = \begin{pmatrix} \hat{f} c_y c_z & \hat{f} c_y s_z \\ \hat{f}(s_x s_y c_z - c_x s_z) & \hat{f}(s_x s_y s_z + c_x c_z) \end{pmatrix}$$

# Step 1: Light Direction

## Camera Calibration

$$H_{2x2} = \begin{pmatrix} \hat{f} c_y c_z & \hat{f} c_y s_z \\ \hat{f}(s_x s_y c_z - c_x s_z) & \hat{f}(s_x s_y s_z + c_x c_z) \end{pmatrix} \quad H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}$$

$$E(\theta_x\, \theta_y\, \theta_z,\, \hat{f}) =$$
$$(\hat{f} c_y c_z - h_1)^2 + (\hat{f} c_y s_z - h2)^2 + (\hat{f}(s_x s_y c_z - c_x s_z) - h4)^2 + (\hat{f}(s_x s_y s_z + c_x c_z) - h5)^2$$

# Step 1: Light Direction

## Camera Calibration

- With an unknown *f*, for estimating $\hat{H}$, we need to find *f* first

$$E(\theta_x\, \theta_y\, \theta_z,\ \hat{f}) =$$
$$(\hat{f}c_y c_z - h_1)^2 + (\hat{f}c_y s_z - h2)^2 + (\hat{f}(s_x s_y c_z - c_x s_z) - h4)^2 + (\hat{f}(s_x s_y s_z + c_x c_z) - h5)^2$$

$$f_1 = \frac{\hat{f}(c_x s_y c_z + s_x s_z)}{h_7} \text{ e } f_2 = \frac{\hat{f}(c_x s_y s_z - s_x c_z)}{h_8}$$

$$f = \frac{h_7^2 f_1 + h_8^2 f_2}{h_7^2 + h_8^2}$$

# Step 1: Light Direction

## Viewer's direction

- In the calibration step, the center of the circle was calculated as C = (C1; C2).
- The center of the limbus in world coordinates is given by Xc = (C1  C2 1)^T.
- In the camera coordinates, the center becomes:

$$x_c = \hat{H} X_c$$

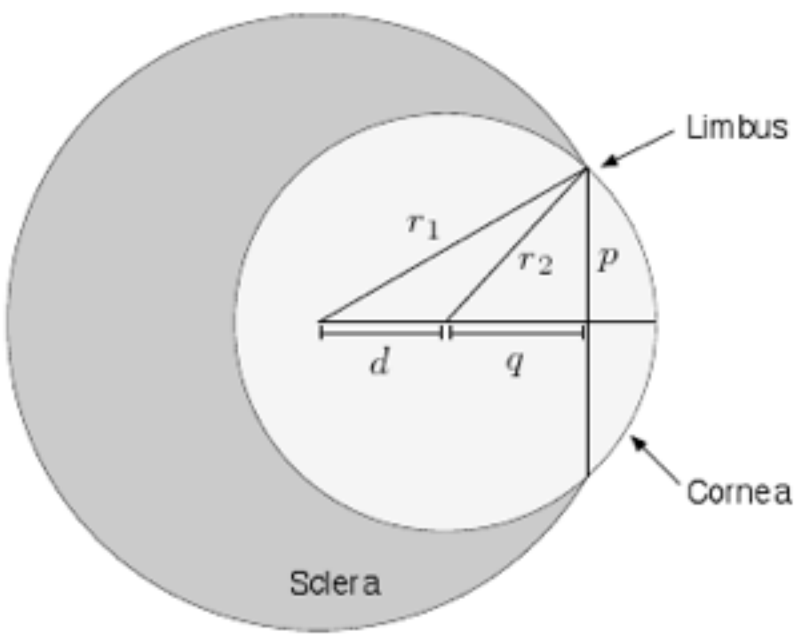
- Therefore, the camera direction is the vector from the center of the limbus and the origin of the camera coordinate system:

$$v = -\frac{x_c}{||x_c||}$$

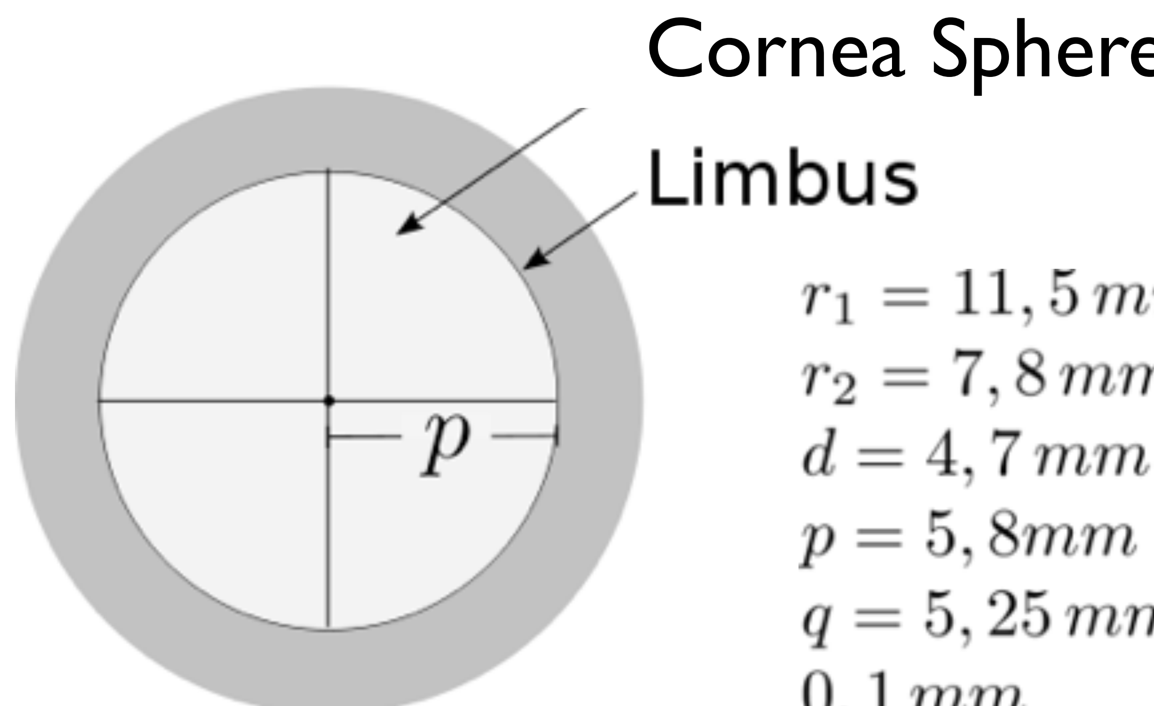
# Step 1: Light Direction

## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye




© Micah 2007

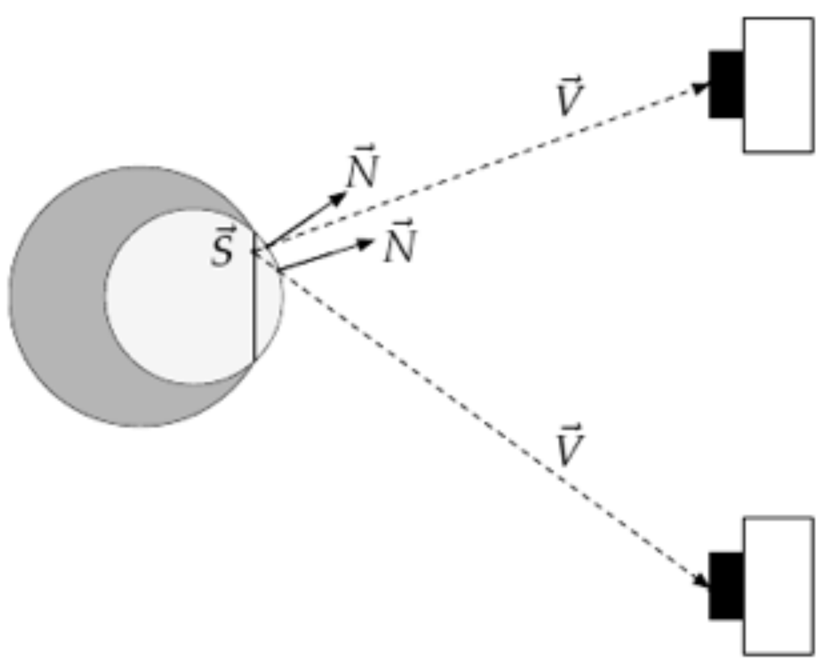



$r_1 = 11,5\,mm$
$r_2 = 7,8\,mm$
$d = 4,7\,mm$
$p = 5,8mm$
$q = 5,25\,mm$
$0,1\,mm$

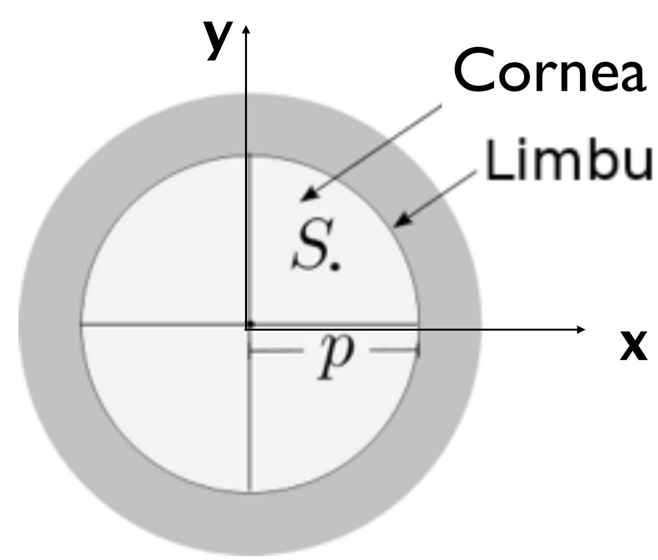
© Priscila & Tiago

# Step 1: Light Direction
## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye



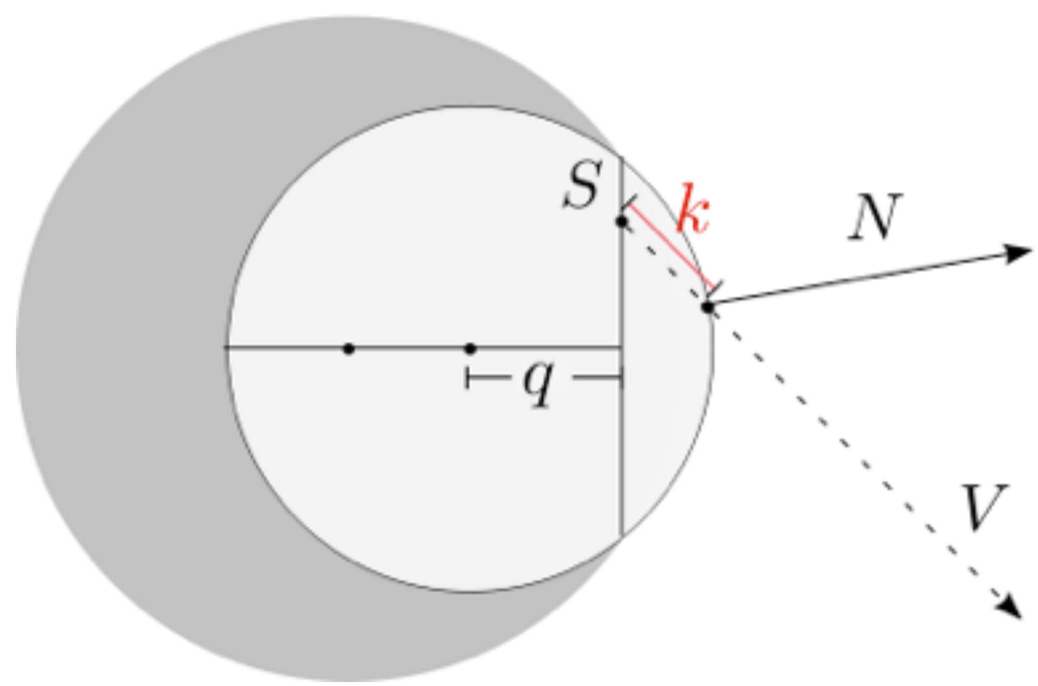

© M. K. Johnson 2007




© Priscila & Tiago

# Step 1: Light Direction
## Surface Normal Direction

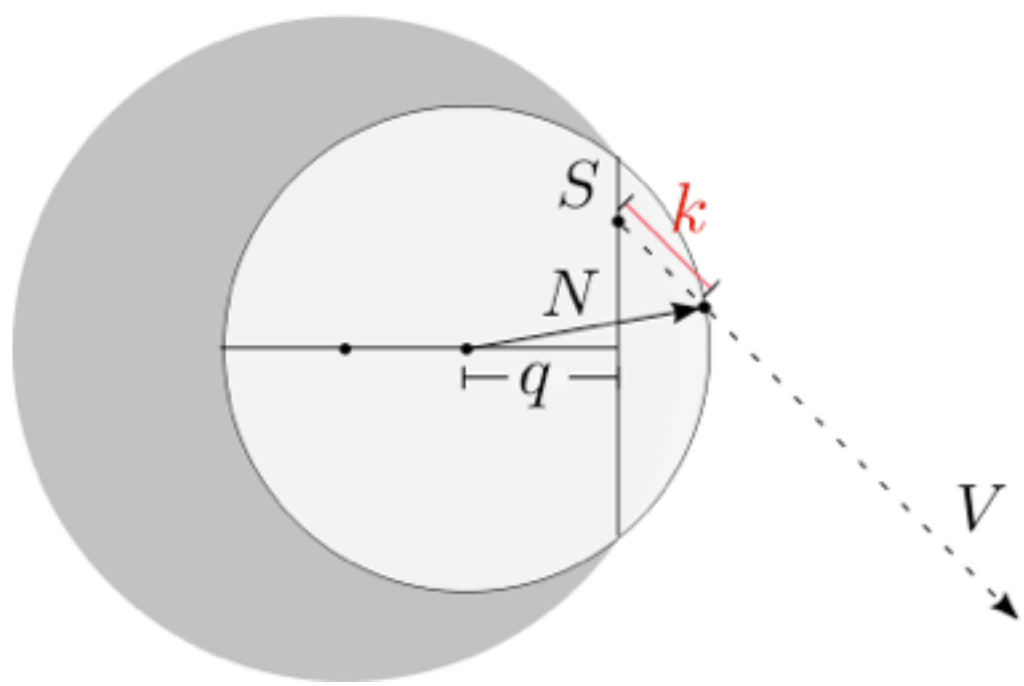
- The 3D surface normal **N** of the eye is estimated from a model of the eye




© Priscila & Tiago

# Step 1: Light Direction
## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye

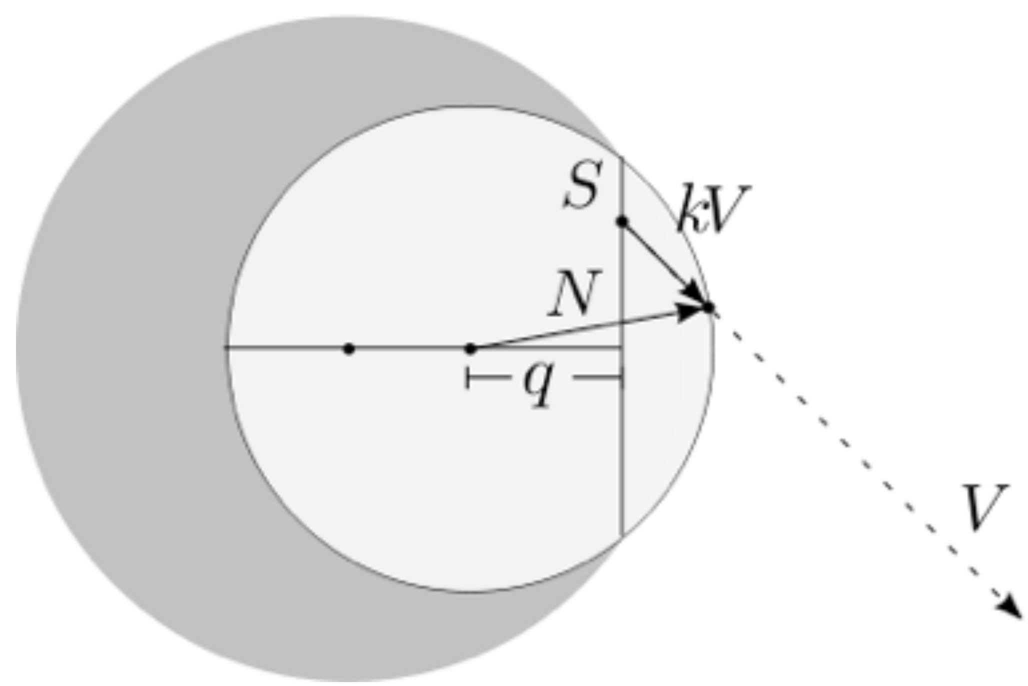



© Priscila & Tiago

# Step 1: Light Direction

## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye

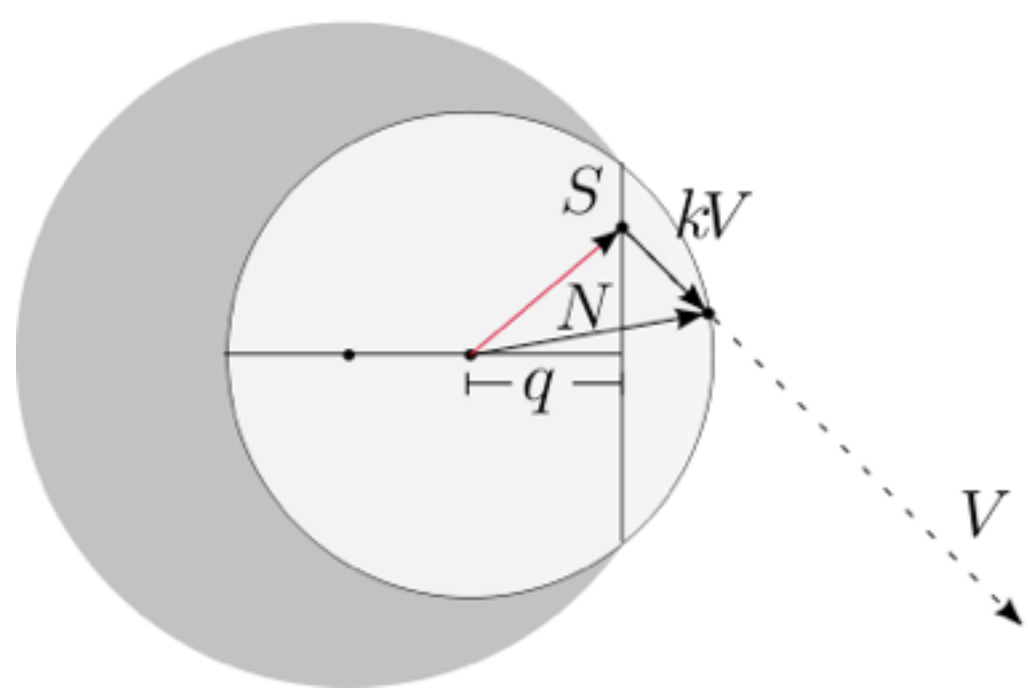



© Priscila & Tiago

# Step 1: Light Direction

## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye

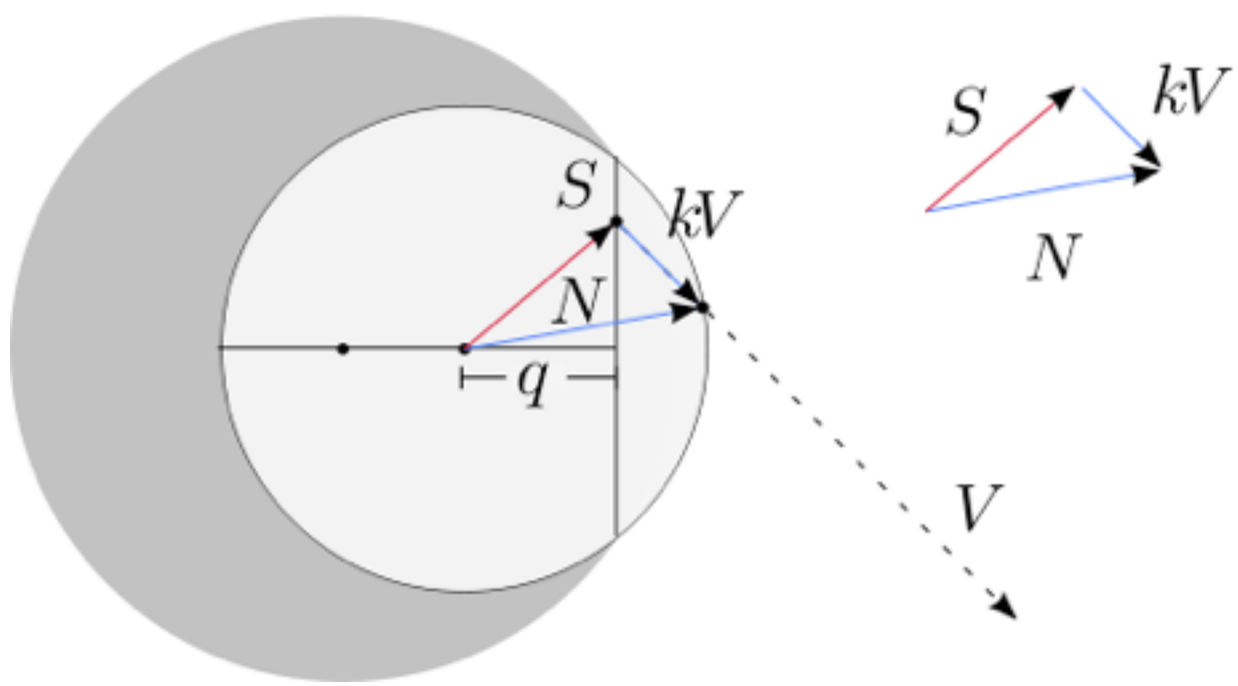



$$S = \begin{pmatrix} S_x \\ S_y \\ q \end{pmatrix}$$

© Priscila & Tiago

# Step 1: Light Direction
## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye

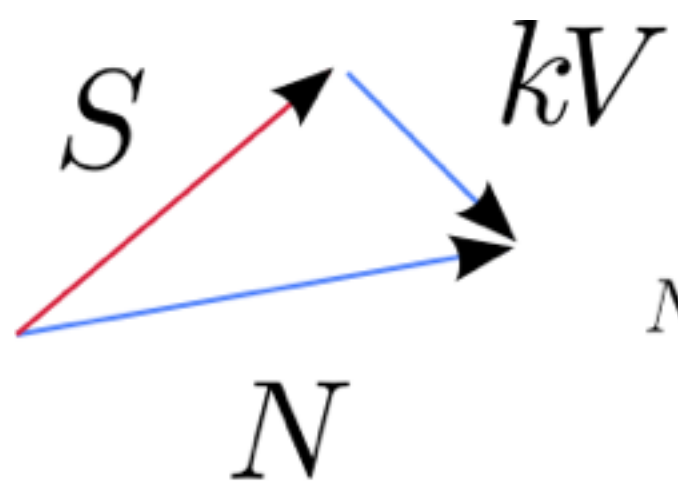



$$N = S + kV$$

© Priscila & Tiago

# Step 1: Light Direction
## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye




$$N = S + kV$$

$$N = \begin{pmatrix} S_x + kV_x \\ S_y + kV_y \\ q + kV_z \end{pmatrix}$$

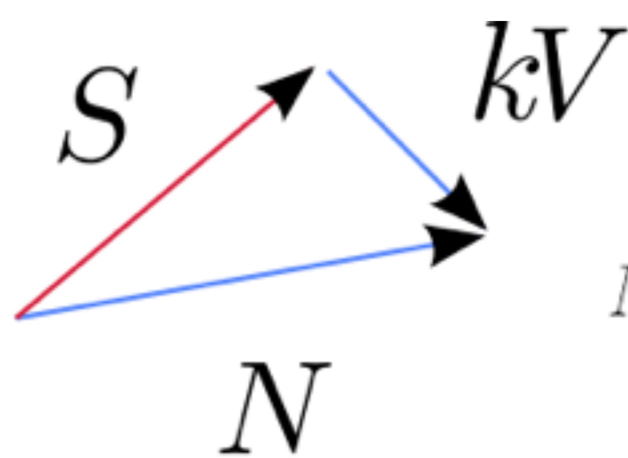
$$X_s = H^{-1} x_s$$

$$S = \frac{q}{r}(X_s - C)$$

# Step 1: Light Direction

## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye

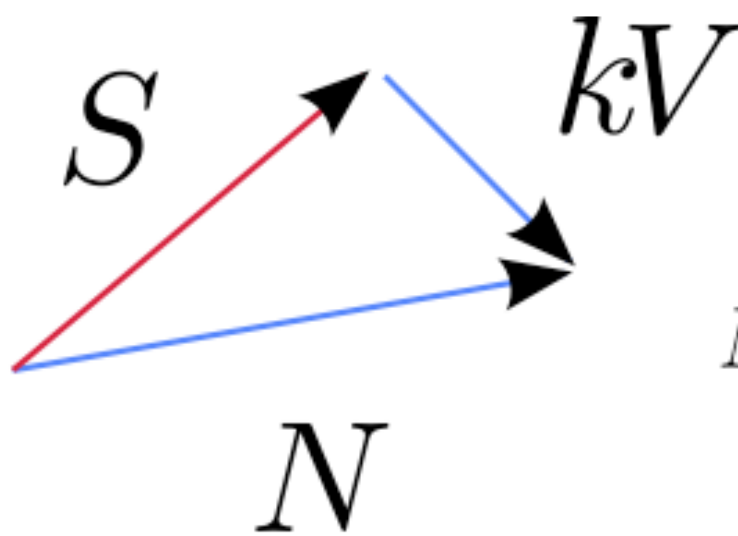



$$N = \begin{pmatrix} S_x + kV_x \\ S_y + kV_y \\ q + kV_z \end{pmatrix}$$

$$V = R^{-1} v$$

# Step 1: Light Direction
## Surface Normal Direction

- The 3D surface normal **N** of the eye is estimated from a model of the eye




$$N = \begin{pmatrix} S_x + kV_x \\ S_y + kV_y \\ q + kV_z \end{pmatrix}$$

$$r_2 = \sqrt{(S_x + kV_x)^2 + (S_y + kV_y)^2 + (q + kV_z)^2}$$

$$k^2 + 2(S_x V_x S_y V_y + qkV_z)k + (S_x^2 + S_y^2 + q^2 - r_2^2 = 0$$

# Step 1: Light Direction

## Light source L

$$L = 2(V^T N)N - V$$

$$l = RL$$

# Step 2:
# Estimating Light Source and Angular Errors

# Consistency check

- Once we estimate the light direction reflected in each eye of a person in the image, it is necessary to check if all of them come from the same light source (i.e., they are consistent with each other)
- The first step is to estimate the direction of the light illuminating the scene
- For that, it is necessary to assume a few conditions

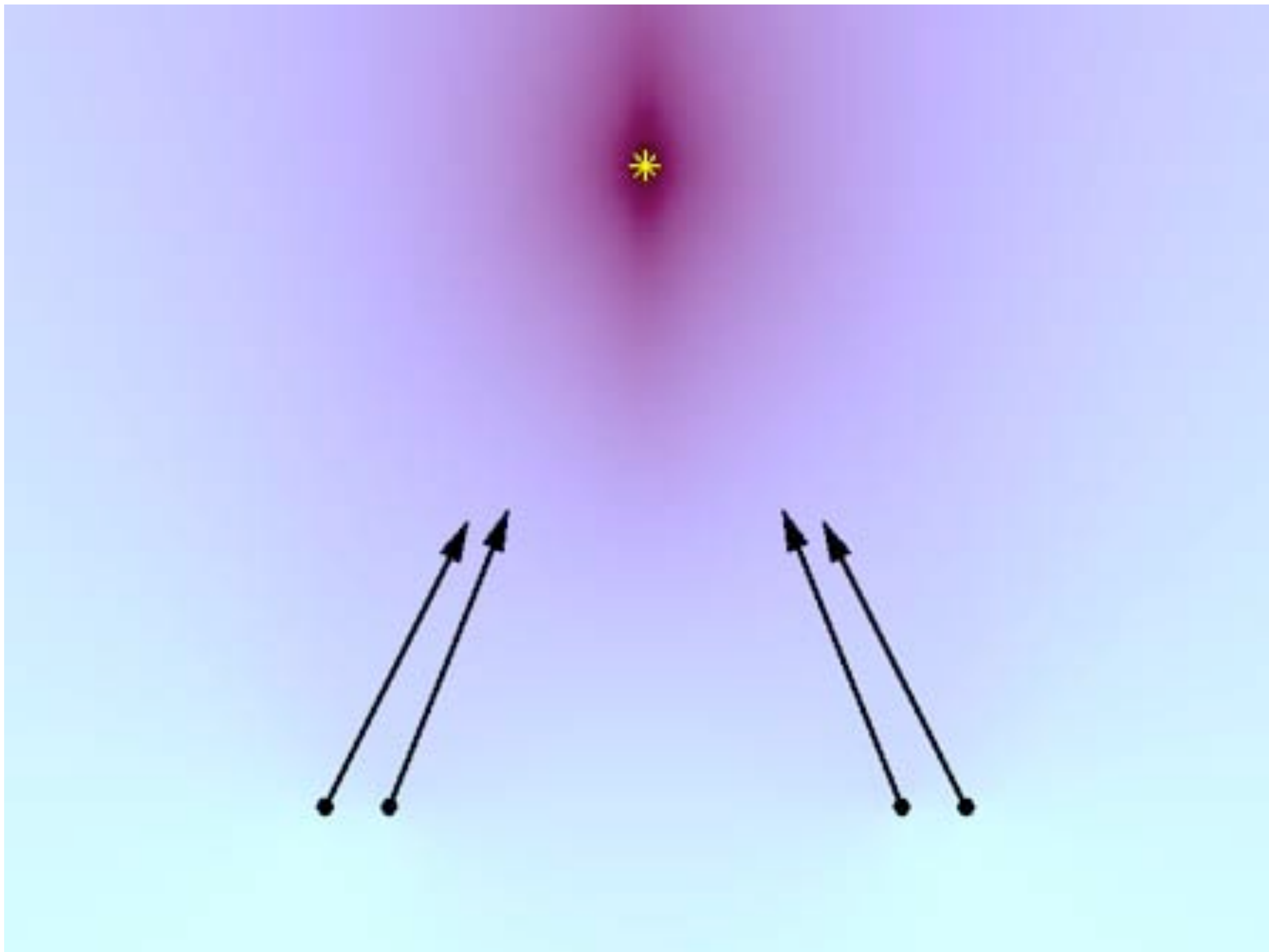
# Consistency check

- Light source are punctual
- Rays converge to the light source
- For each eye $i$, the angle between the vector to the light source in the position x and the estimated direction is

$$\theta_i(\vec{x}) \;=\; \cos^{-1}\left(\vec{l}_i^{\,T}\,\frac{\vec{x}-\vec{p}_i}{\|\vec{x}-\vec{p}_i\|}\right)$$

- $\vec{p}_i$ is the position of the i-th light ray

# Consistency check



© Micah 2007

# Consistency check

- Given N light rays and their respective estimated light source positions, the punctual source light can be estimated by maximizing the error function bellow using a non-linear conjugate gradient technique

$$\hat{E}(\vec{x}) = \sum_{i=1}^{N} \vec{l}_i^T \frac{\vec{x} - \vec{p}_i}{\|\vec{x} - \vec{p}_i\|}$$

# Consistency check

- Therefore, $\vec{x}^*$ denotes the light source position
- The angle between the i-th light ray and the light source is given by

$$\theta_i(\vec{x}^*)$$

- After estimating the scene light source, we need to decide whether or not the image is a product of a composition

# Experiments and Validation

# Experiments

**Database of images:**

- 120 images: 60 authentic + 60 composed;
- Images with resolution bigger than 3 megapixels;
- Images containing two or more people, with clearly visible eyes.

# Experiments

Database of images:

Pristine

Fake

# Experiments

- The pixels of the limbus, as well as the specular highlights are manually marked;
- For each image, the four LME, LSE, VME and VSE features are appropriately extracted;
- Each data set is composed by 120 feature vectors, since we have 120 images (5 sets of extractions).

# Experiments

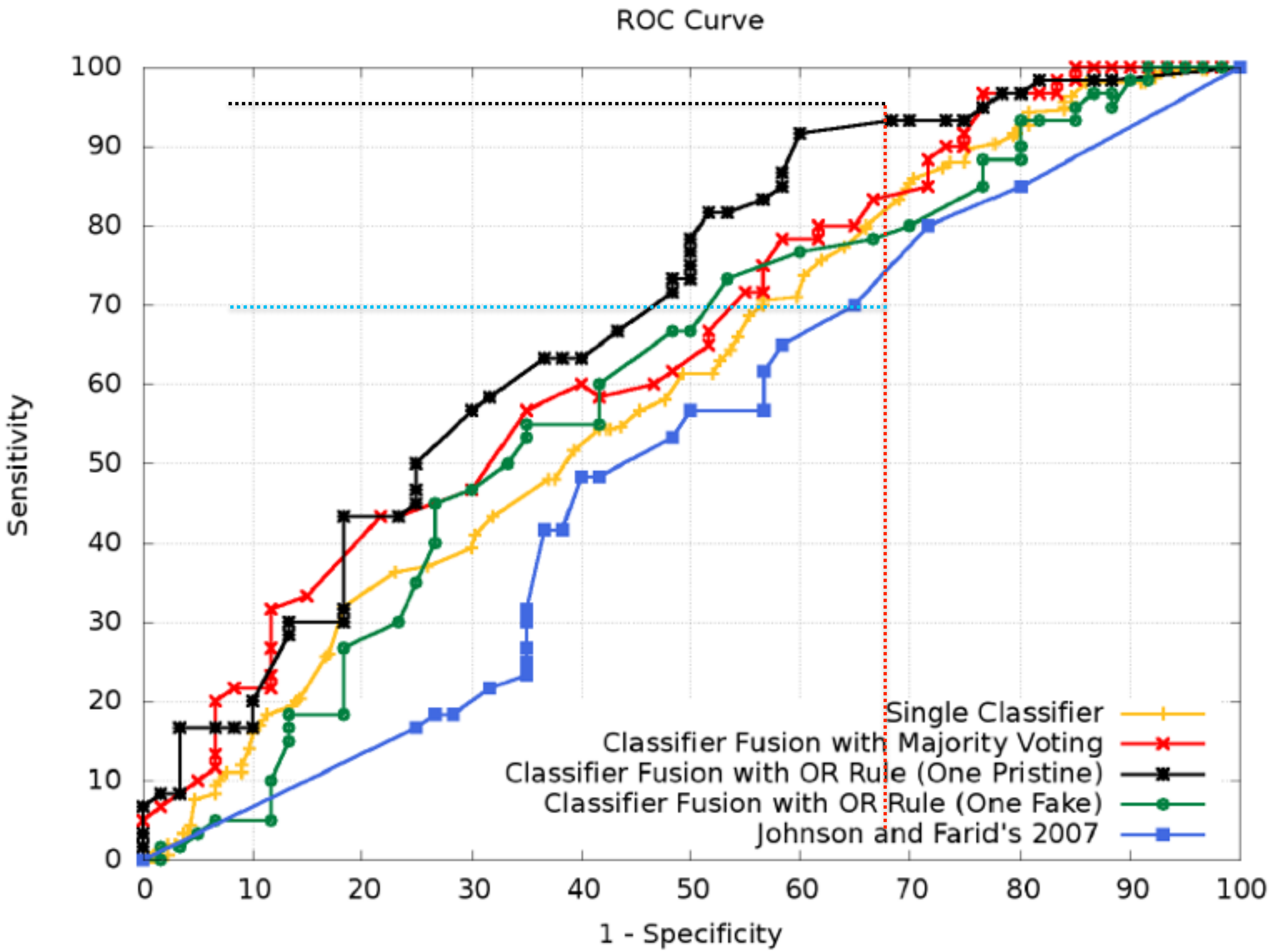
- Two-class classifiers (Out-of-box SVM with RBF kernel).
- We performed 5-fold cross validation in all experiments.
- Results for a single classifier and for three pools of classifiers.
- Each pool is composed by 5 classifiers, one for each data set.
- Process decision making: based on a classifier fusion fashion approach.

# Results

Applied classification strategies:

- Single Classifier *(SC)*
- Classifier Fusion with Majority Voting *(MV)*
- Classifier Fusion with OR Rule *(One Pristine)*
- Classifier Fusion with OR Rule *(One Fake)*

# Results



ROC Curve
Sensitivity
1 - Specificity
0
10
20
30
40
50
60
70
80
90
100
Single Classifier
Classifier Fusion with Majority Voting
Classifier Fusion with OR Rule (One Pristine)
Classifier Fusion with OR Rule (One Fake)
Johnson and Farid's 2007

# Results (EER)

| | EER (%) | Accuracy(%) | Improv. over prior work (%) |
|---|---|---|---|
| Single Classifier | 44 | 56 | 7 |
| Fussion MV | 40 | 60 | 15 |
| **Fussion One Pristine** | **37** | **63** | **21** |
| Fusion One Fake | 41 | 59 | 13 |
| Johnson and Farid's | 48 | 52 | - |

# Conclusions

- Our proposed extension has potential to pinpoint composites of people in photographs, as Johnson and Farid's work.
- The extension uses more discriminative features that improves the detection of tampering.

# Conclusions

- The new features and the new decision-making processes reduce the classification error in more than 20% when compared to prior.