Camera Parameters and Calibration

Paulo Dias

Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Definitions - Luminance

Luminance

Luminance is normally defined as a measurement of the photometric luminous intensity per unit area of light travelling in a given direction.

Therefore it is used to describe the amount of light that goes through, or is emitted from, a particular area, and falls within a given solid angle.

The SI unit for luminance is candela per square meter (cd/m2).

The CGS unit of luminance is the *stilb*, which is equal to one candela per square centimeter or 10 kcd/m2.

Definitions - Chrominance

Chrominance

Chrominance is a numeral that describes the way a certain amount of light is distributed among the visible spectrum.

A black and white image has a balanced distribution of energy among to the visible spectrum matched to the band pass characteristics of the human visual system. This means that when viewed by a human a B&W image has no color information which means that its color information is zero.

Therefore, chrominance has no luminance information but is used together with it to describe a colored image defined, for instance, by an RGB triplet.

Any RGB triplet in which the value of R=G=B has no chrominance information.

Separating Luminance from Chrominance

Given an RGB triplet, we can define a derived triplet in which luminance and chrominance can be separated:

Luminance

$$Y = W_r R + W_g G + W_b B$$

$$U = U\max\frac{B-Y}{1-W_b} \approx 0.492(B-Y)$$

$$V = V\max\frac{R-Y}{1-W_r} \approx 0.877(R-Y)$$

Chrominance

where

$$W_r = 0.299$$
$$W_B = 0.114$$
$$W_G = 0.587$$
$$U_{\max} = 0.436$$
$$V_{\max} = 0.615$$

This values originally derivates from the general model of the human visual system and had a significant impact on the ability to develop a television color system compatible with the previous B&W television systems.

A symetric operation can be performed in order to recover the original RGB triple.

Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Image processing pipeline

A typical image processing pipeline (inside the image device) for a tri-stimulus system is shown bellow. This processing can be performed on the YUV or RGB components depending on the system. This should be understood as a mere example.

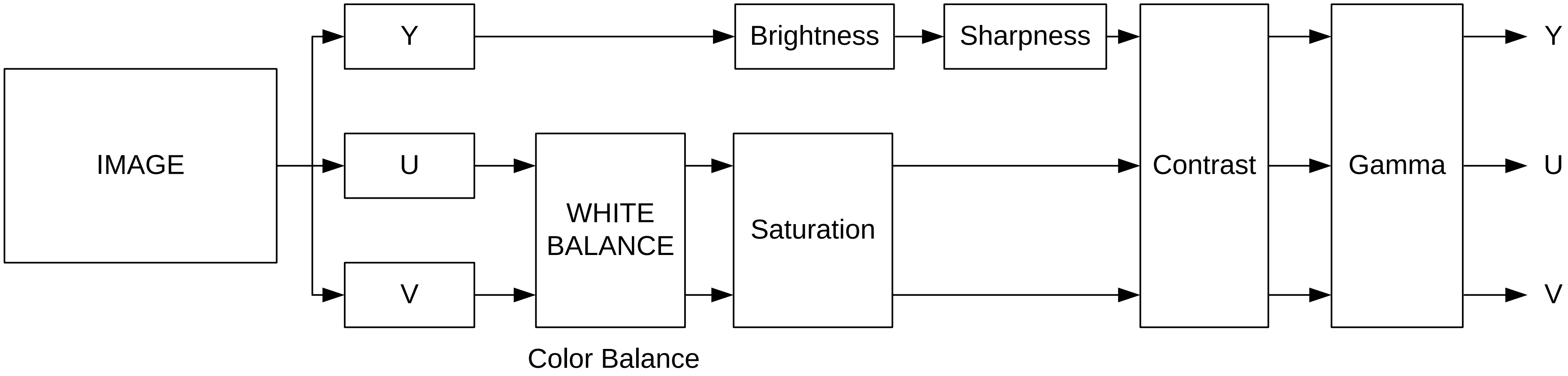
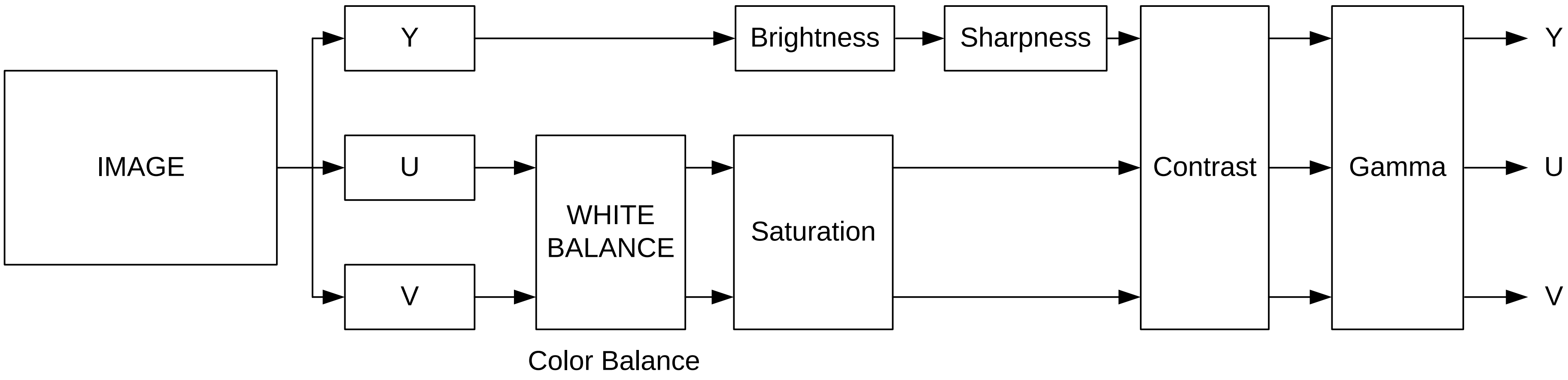


Image processing pipeline

Depending on the system, more or less image parameters may be available for the user to control. Also, some of these parameters (namely brightness, contrast and saturation) are also intrinsic original image characteristics apart from being externally controllable parameters.



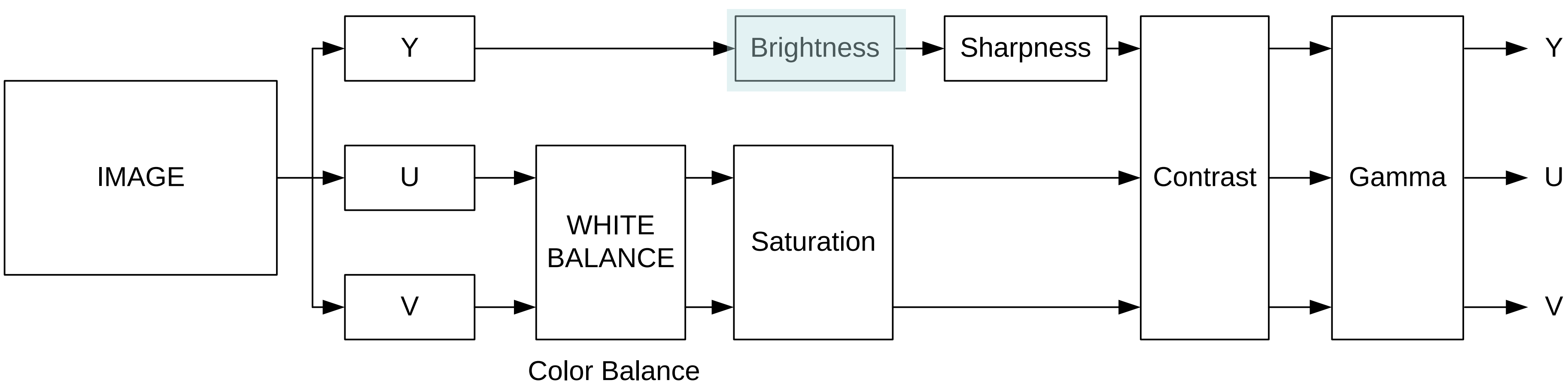
Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Brightness (as an intrinsic image characteristic)

Brightness is one on the intrinsic original image characteristics. It represents a measure of the average amount of light that is integrated over the image during the exposure time. Exposure time (that is, the period of time during which the sensor receives light while forming the image, may or may not be a controllable parameter of the image device).

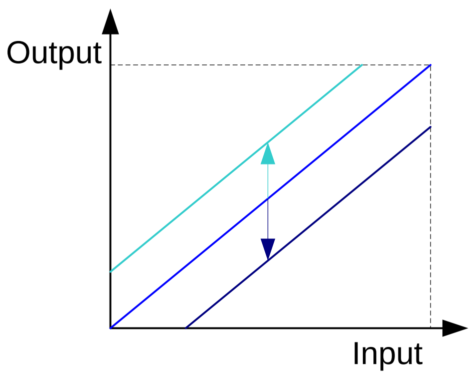
If the brightness it too high overexposure may occur which will white saturate part or the totality of the image.

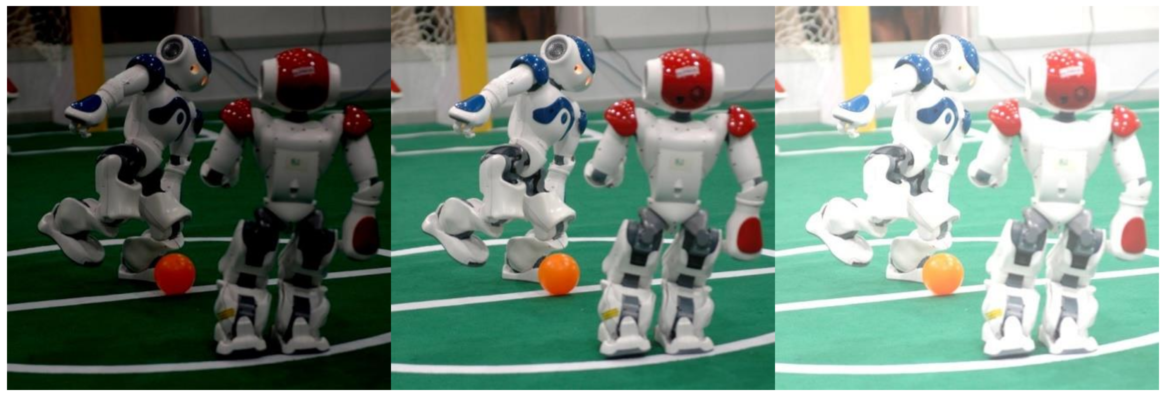


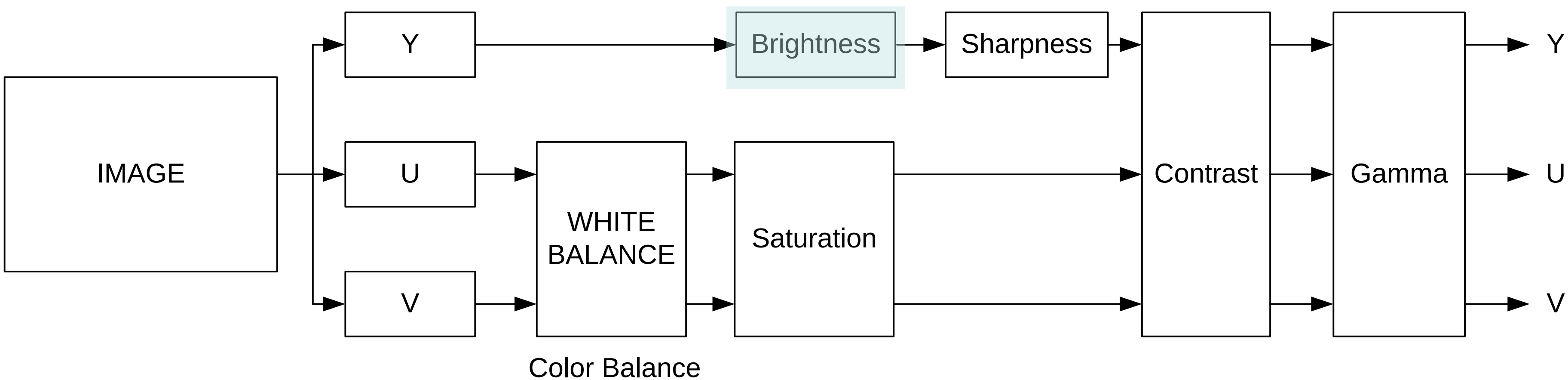
Brightness

Brightness (as a controllable parameter)

The brightness parameter is basically a constant (or offset) that can be added (subtracted) from the luminance component of the image.



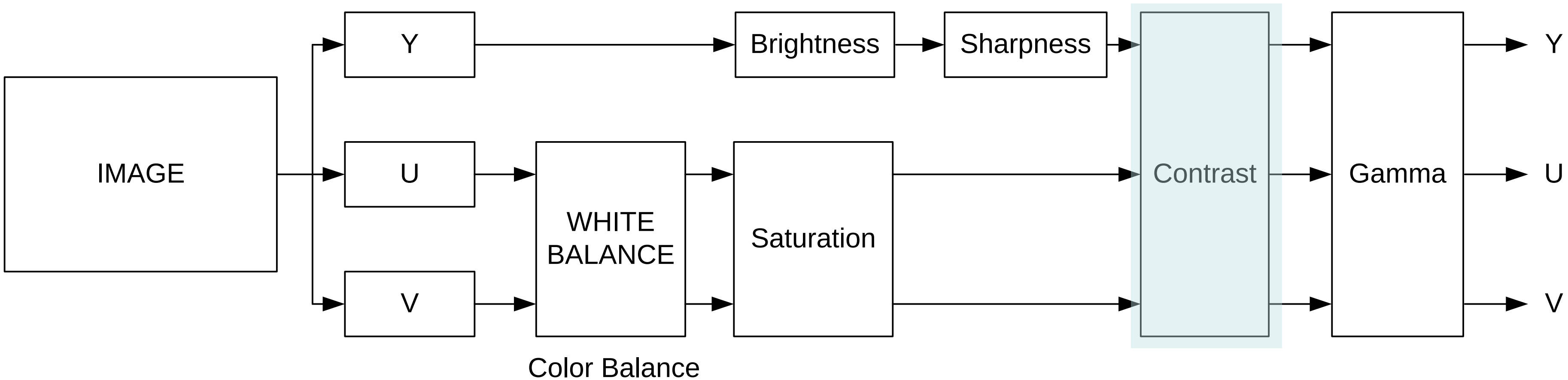




Contrast (as an intrinsic image characteristic)

There is not a unique definition of contrast. On of the most used is that contrast is the difference in luminance (or color) along the 2D space that makes an object distinguishable. In visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view. The faster and higher the luminance (or color) changes along the space the higher the contrast is.

The maximum possible contrast of an image is also denominated contrast ratio or dynamic range.

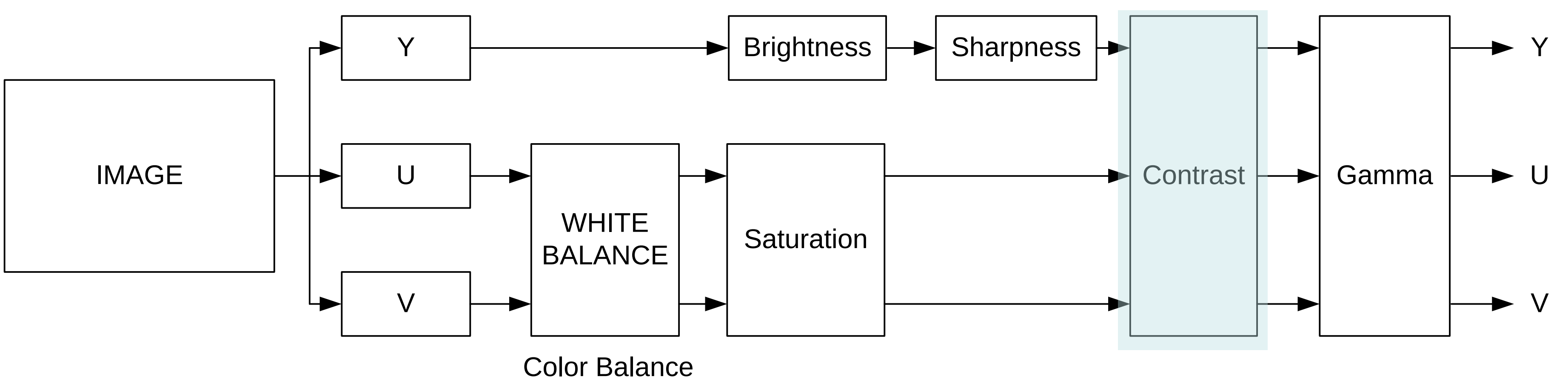


Contrast (as an intrinsic image characteristic)

One of the possible definitions of contrast is given by the expression

$$\frac{\text{Luminance diference}}{\text{Average luminance}}$$

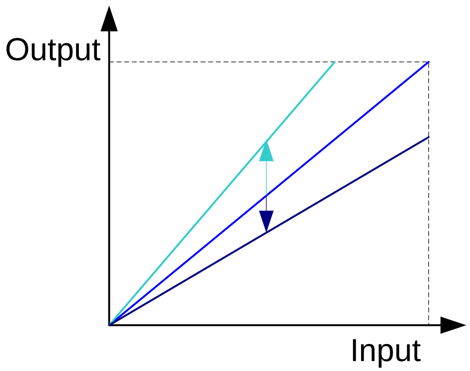
The human eye contrast sensitivity function is a typical band-pass filter with a maximum at around 4 cycles per degree with sensitivity reducing to both sides off that maximum. This means that the human visual system can detect lower contrast differences at 4 cycles per degree than at any other spatial frequency.

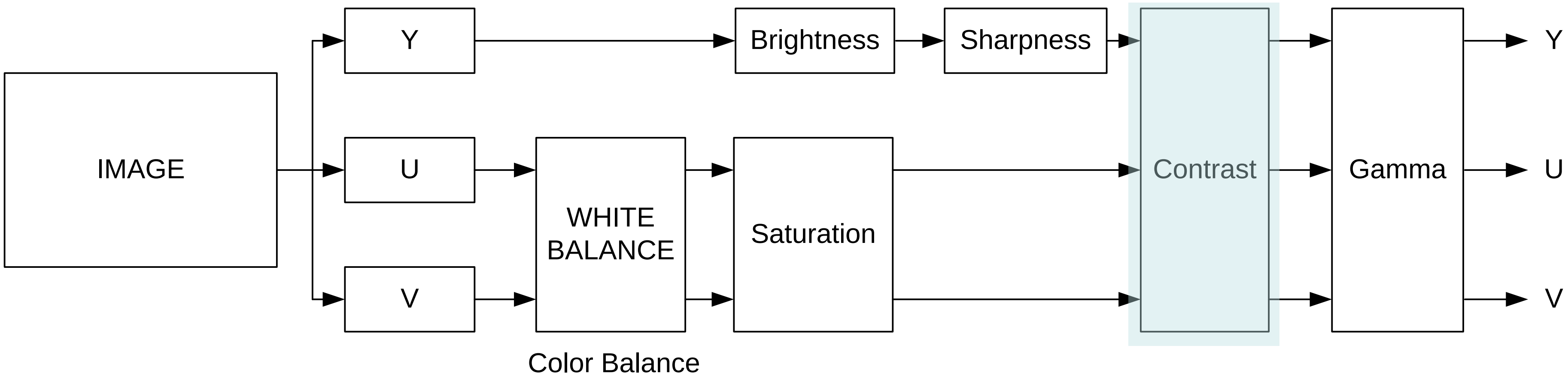


Contrast

Contrast (as a controllable parameter)

The contrast parameter is basically a variation in the gain control function of the luminance component of the image.

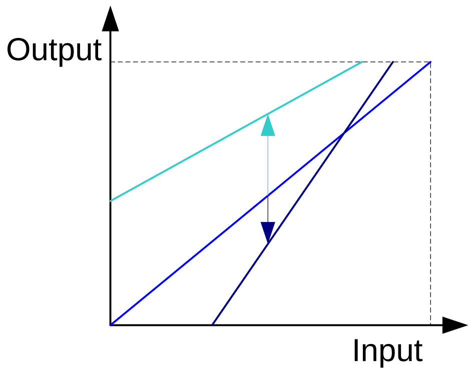


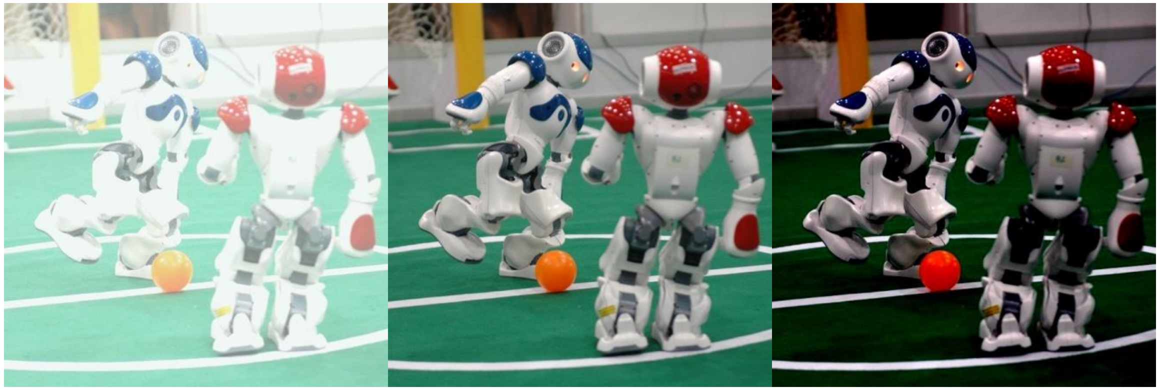


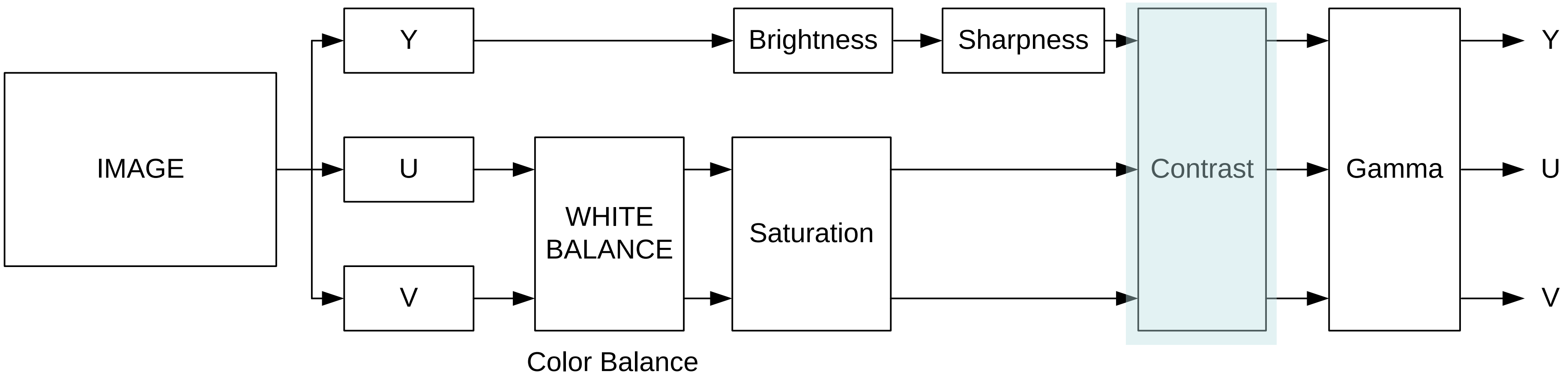
Contrast + Brightness

Contrast + Brightness(as controllable parameters)

It is common that contrast and brightness are actually a combined single transfer function.





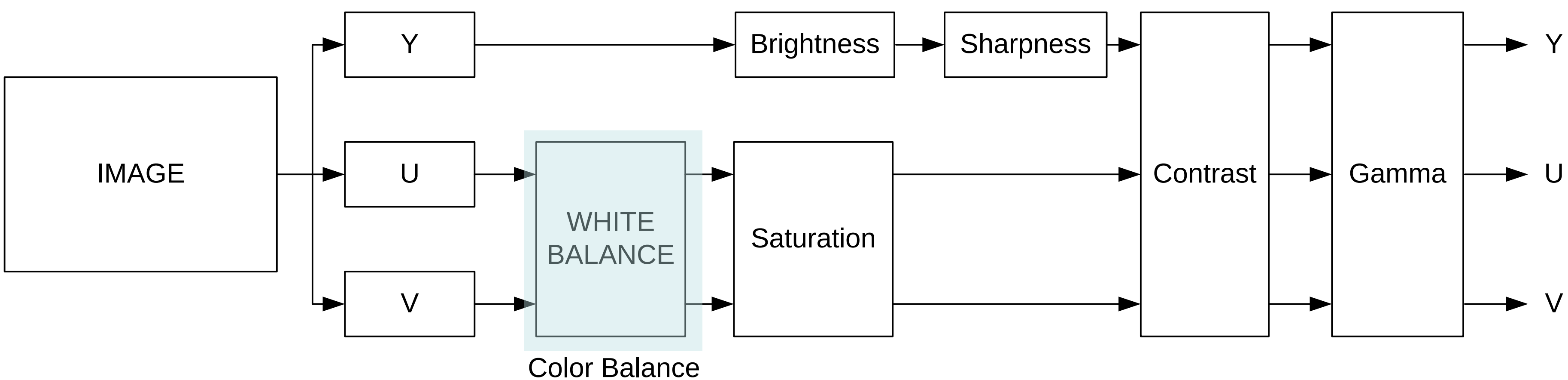


White Balance(as controllable parameters)

White balance is the global adjustment of the intensities of the colors (typically red, green, and blue primary colors).

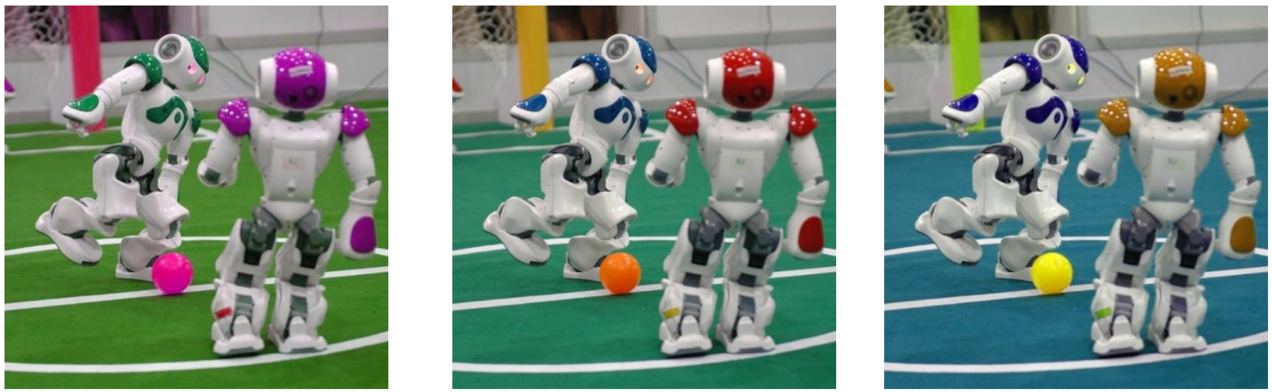
An important goal of this adjustment is to render specific colors – particularly neutral colors – correctly; hence, the general method is sometimes called gray balance, neutral balance, or white balance.

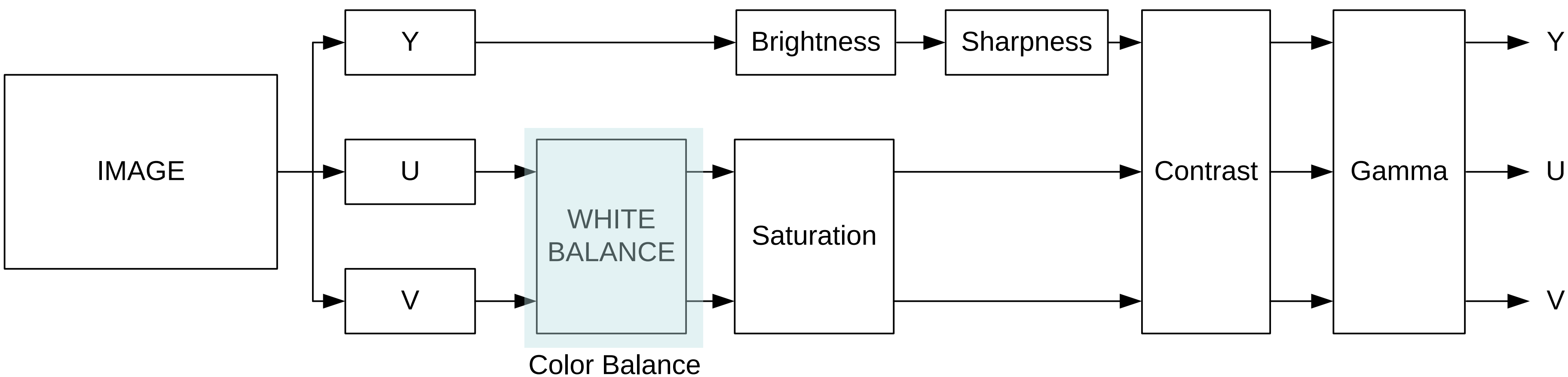
This balance is required because of different color spectrum energy distribution depending on the illumination source.



White Balance

Examples





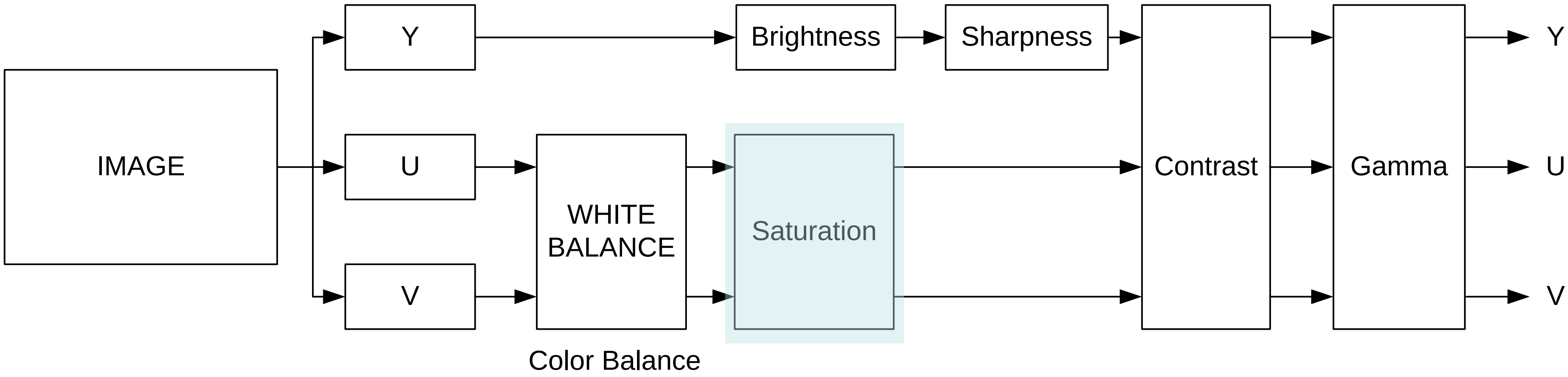
Saturation

Saturation (as an intrinsic image characteristic)

The saturation of a color is determined by a combination of light intensity that is acquired by a pixel and how much this light it is distributed across the spectrum of different wavelengths. The most purest (most saturated) color is obtained when using a single wavelength at a high intensity (laser light is a good example).

If the light intensity declines, then, as a result, the saturation also decline.

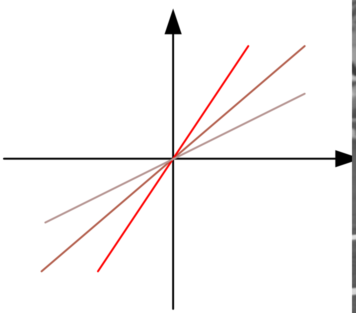
A non saturated image (B&W) has a spectrum distribution that matches the human eye spectrum sensibility. Saturation is sometimes also defined as the amount of white you have blended into a pure color.



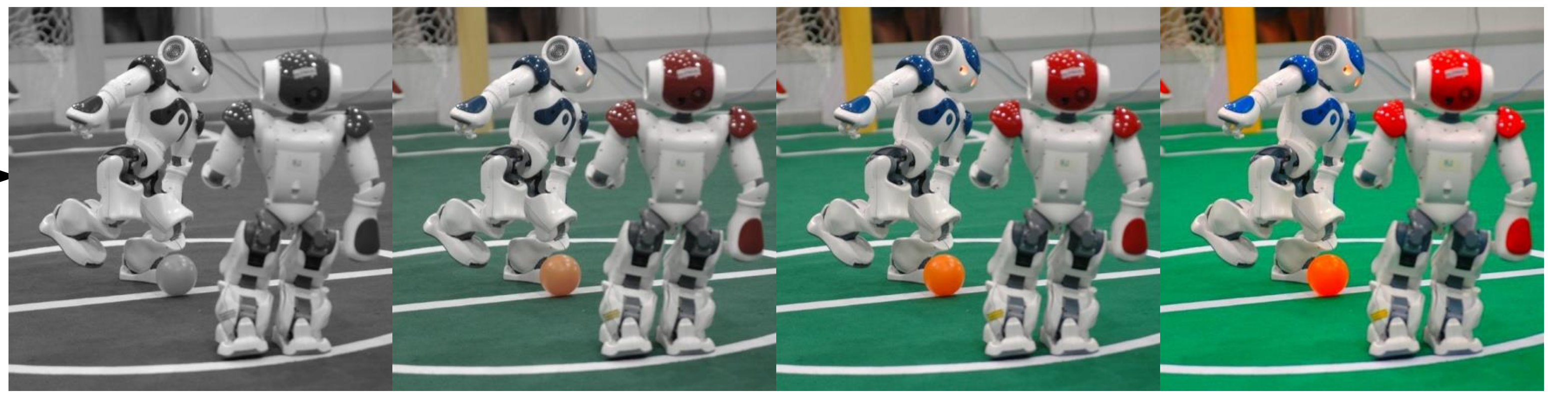
Saturation

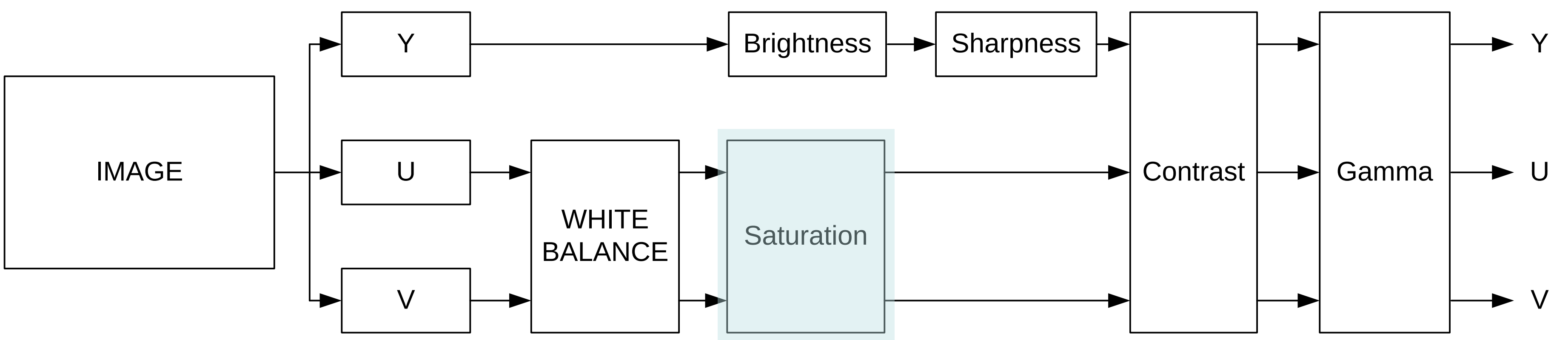
Saturation (as a controllable parameter)

To reduce the saturation of an image we can add white to the original colors. In fact this is the same as changing the gain of the U and V chromatic components.



UV transfer function





Color Balance

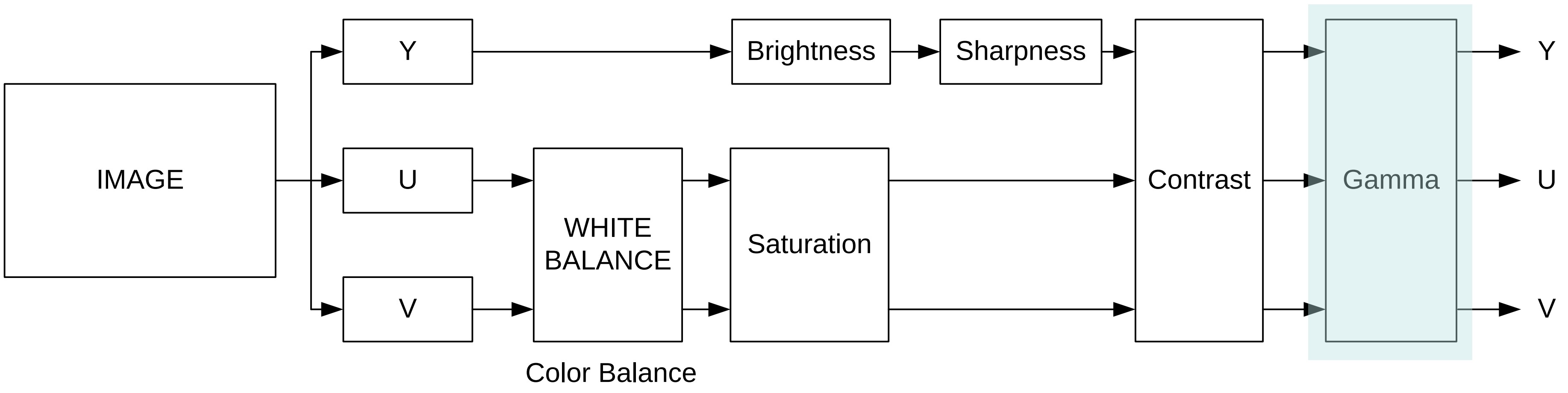
Gamma

Gamma correction is the name of a nonlinear operation used to code and decode luminance or TGB tristimulus values. In the simplest cases gamma is defined by the power-law expression:

$$V_{out} = AV_{in}^{\delta}$$

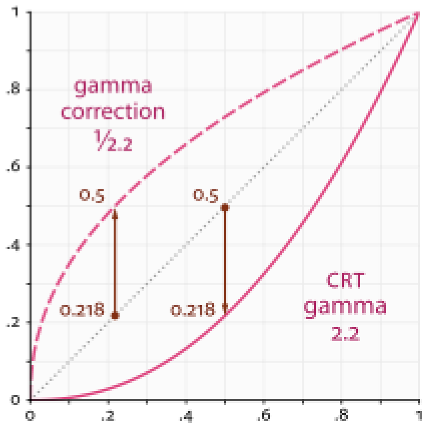
where *A* is a constant and the input and output values are non-negative real values.

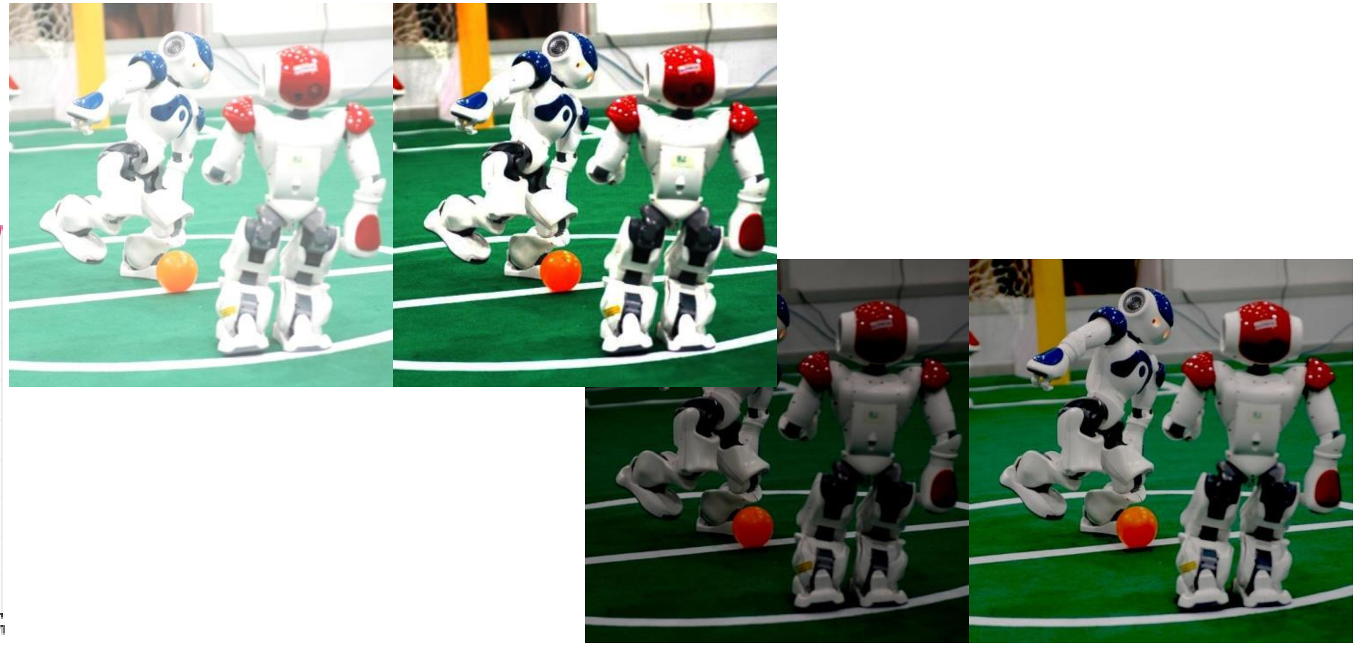
In most cases A = 1, and inputs and outputs are typically in the range 0–1.

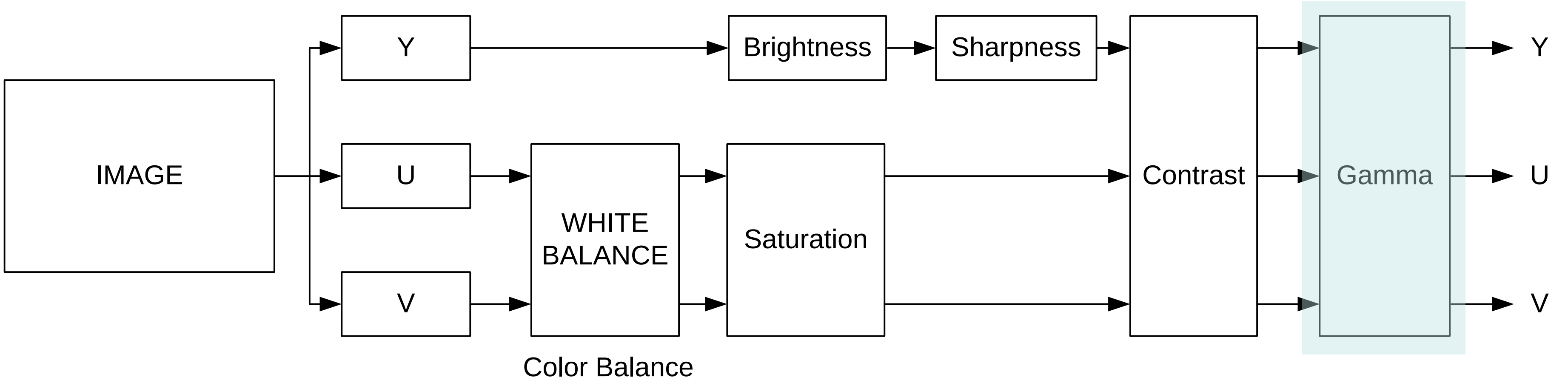


Gamma

Examples







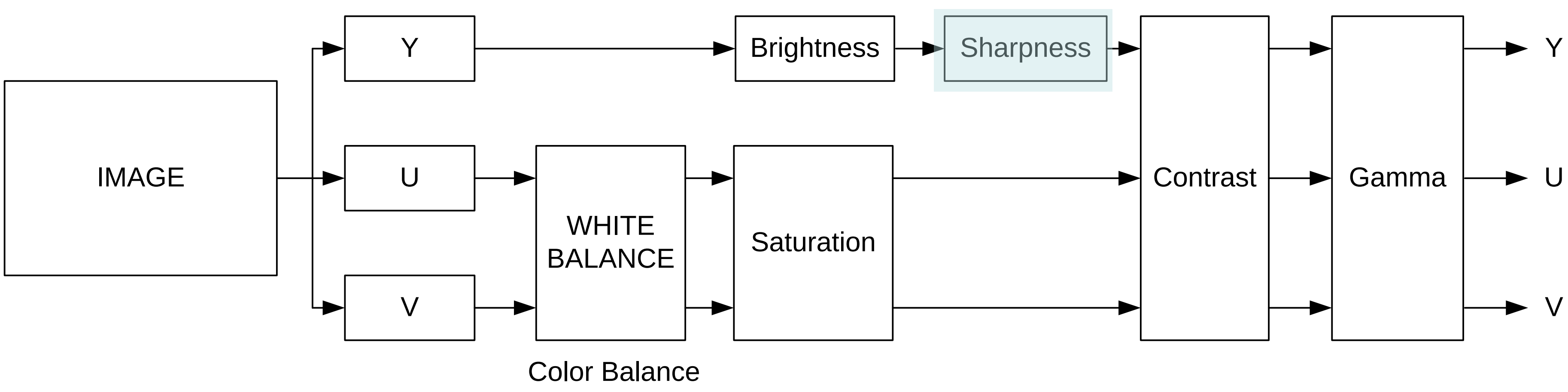
Sharpness

Sharpness (as a controllable parameter)

Sharpness is a measure of the energy frequency spatial distribution over the image.

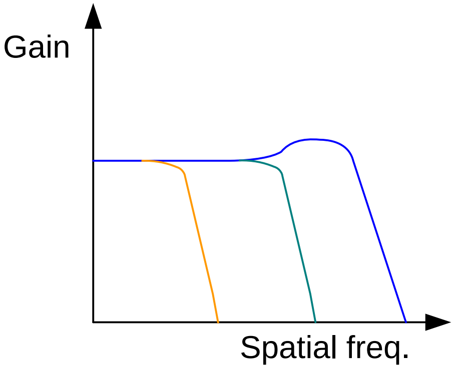
Not all devices provide access to this parameter.

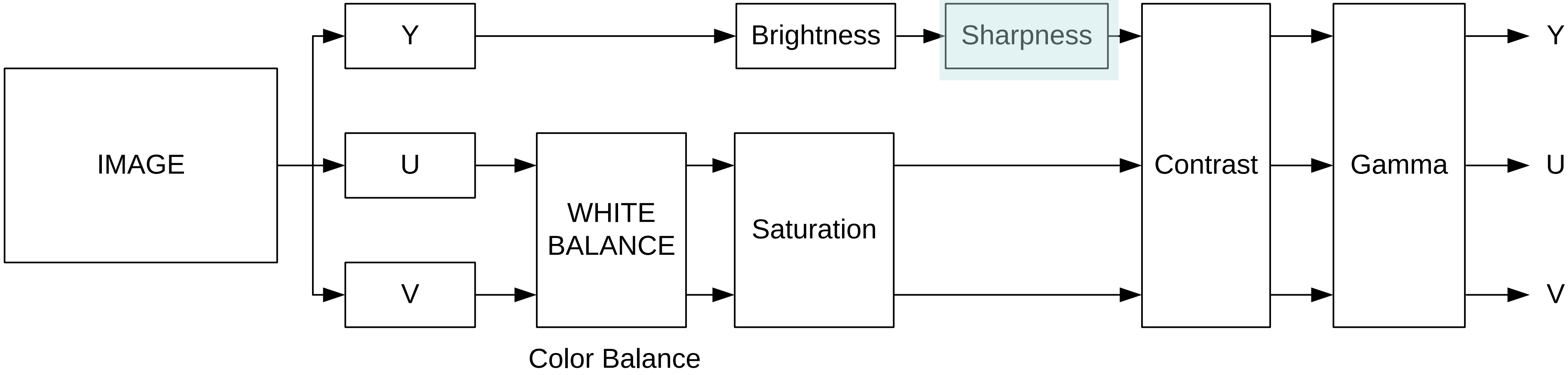
Sharpness basically allows the control of the cut-off frequency of a low pass spatial filter. This may be very useful if the image is afterward intended to be decimated, since it allows to prevent spatial aliases artifacts.



Sharpness (as a controllable parameter)

Examples.





Sumary

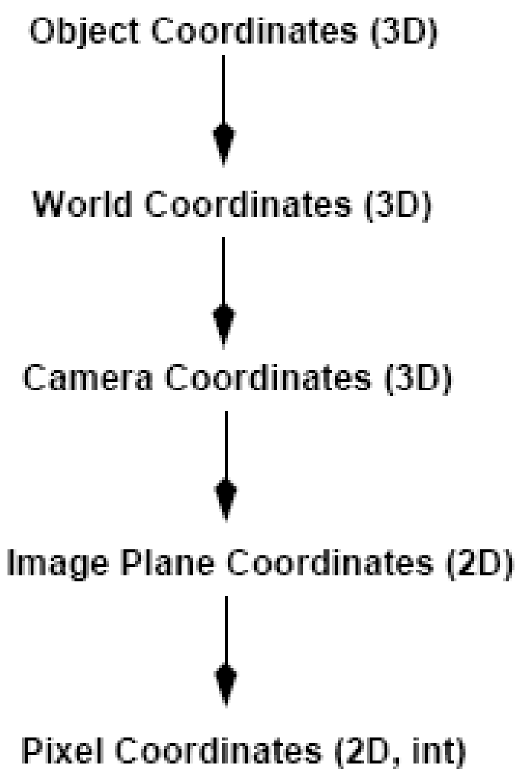
- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Camera parameters

One of the many applications of computer vision is to extract, from the frame image, information regarding the external world from which it has a limited perspective.

These applications may include, for instance, the determination of a certain distance of a point represented in the image when evaluated in real world coordinate system. This is only valid if we also know the plane in which this original point lies.

Therefore, geometrical transformations must be performed in order to accommodate the camera internal parameters and geometry and its position and posture in the real world.



Camera parameters

Camera parameters are normally divided into two big groups:

<u>Extrinsic:</u> the parameters that define the *location* and *orientation* of the camera reference frame with respect to a known world reference frame.

<u>Intrinsic:</u> the parameters necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame.

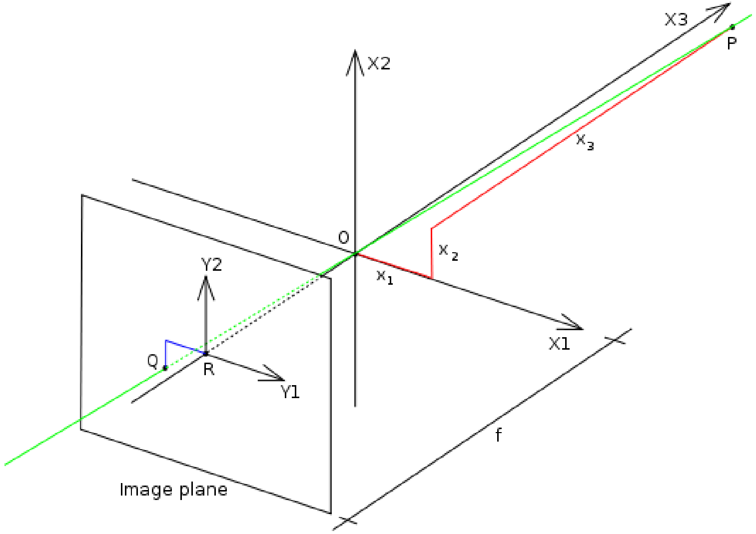
The pinhole model

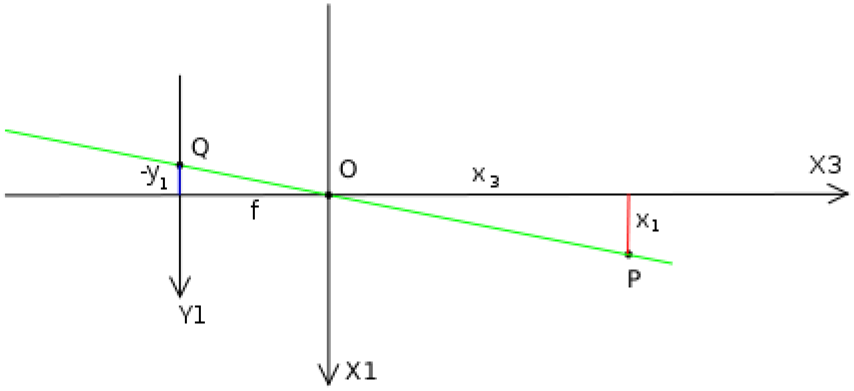
For most practical digital image processing tasks, the optical system can be approximated by the pinhole model.

In this model, the lens is replaced by a very narrow opening (pinhole) through which lights go through directly into the image acquisition plane.

The point that is stroked by a light ray going through the pin hole in the direction of the lens main axis is called the image plane origin

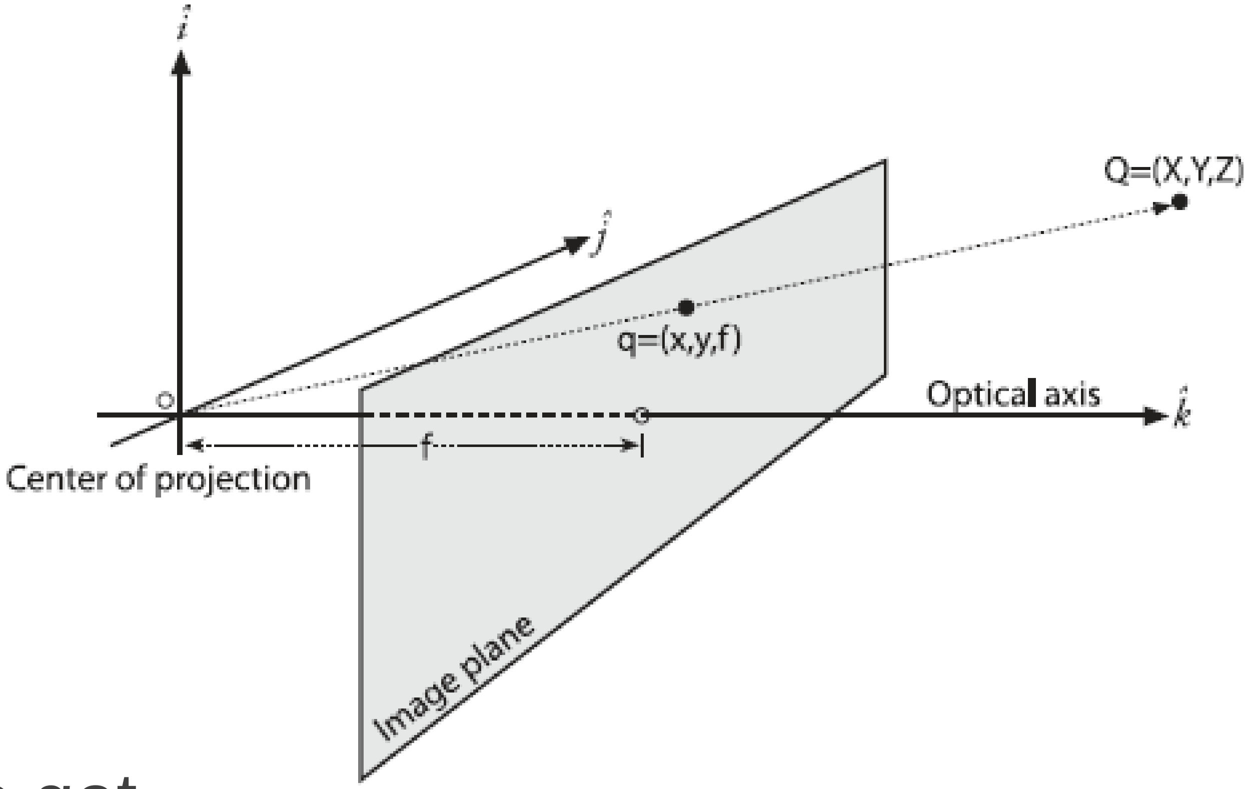
In this model the acquisition plane lies at the focal distance from the pinhole.





Camera model

- Pinhole camera model



- We get

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$

Camera model: optical centre

- Image optical centre and sensor centre are normaly not aligned:

$$x_{screen} = f_x \frac{X}{Z} + c_x$$
$$y_{screen} = f_y \frac{Y}{Z} + c_y$$

- f_x e f_y are derived from focal distance but take in consideration pixel size of the sensor (typically rectangular).

Homogenous Coordinates

Homogenous Coordinates, in comparison to Cartesian coordinates, add an extra coordinate and define an equivalence relationship

(x,y) -> (kx, ky, k)

(X,Y,Z) -> (wX, wY, wZ, w)

This implies that any point in a 3D space can be represented by a multitude of equivalent matrixes, since the cartesian coordinates can be recovered from the homogenous coordinates by dividing each coordinate by the factor *w.*

In fact, this even allows us to represent any point in a plane which is at an infinite distance from the origin. Such representation will have its $w = 0$

Homogenous Coordinates

A rotation of a vector defined by two points in a Cartesian system can be obtained from (point a_n as the coordinates of the point to be projected, c_n as the pinhole coordinates and d_n as resulting rotated vector.

$$\begin{bmatrix}\mathbf{d}_x\\ \mathbf{d}_y\\ \mathbf{d}_z\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta_x & -\sin\theta_x\\ 0 & \sin\theta_x & \cos\theta_x\end{bmatrix}\begin{bmatrix}\cos\theta_y & 0 & \sin\theta_y\\ 0 & 1 & 0\\ -\sin\theta_y & 0 & \cos\theta_y\end{bmatrix}\begin{bmatrix}\cos\theta_z & -\sin\theta_z & 0\\ \sin\theta_z & \cos\theta_z & 0\\ 0 & 0 & 1\end{bmatrix}\left(\begin{bmatrix}\mathbf{a}_x\\ \mathbf{a}_y\\ \mathbf{a}_z\end{bmatrix} - \begin{bmatrix}\mathbf{c}_x\\ \mathbf{c}_y\\ \mathbf{c}_z\end{bmatrix}\right)$$

The translation vector ***T_v***, for the point ***p*** can, on the other hand, be defined in a homogenous form by

$$T_{\mathbf{v}}\mathbf{p} = \begin{bmatrix}1 & 0 & 0 & v_x\\ 0 & 1 & 0 & v_y\\ 0 & 0 & 1 & v_z\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}p_x\\ p_y\\ p_z\\ 1\end{bmatrix} = \begin{bmatrix}p_x + v_x\\ p_y + v_y\\ p_z + v_z\\ 1\end{bmatrix} = \mathbf{p} + \mathbf{v}$$

Homogenous Coordinates

One of the most interesting things in the use of homogenous coordinates is that allows to combine a rotation matrix and a translation matrix into a single homogeneous matrix.

Rotation Matrix

Translation Matrix

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} - & - & - & - \\ - & R & - & -RT \\ - & - & - & - \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Homogeneous coordinates

$$x_{screen} = f_x \frac{X}{Z} + c_x$$
$$y_{screen} = f_y \frac{Y}{Z} + c_y$$

- In homogeneous coordinates, previous equations can be written as
- $q = MQ$

with

- $q = \begin{bmatrix} x \\ y \\ w \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$
- Normalizing with w=1, we get the same equations.

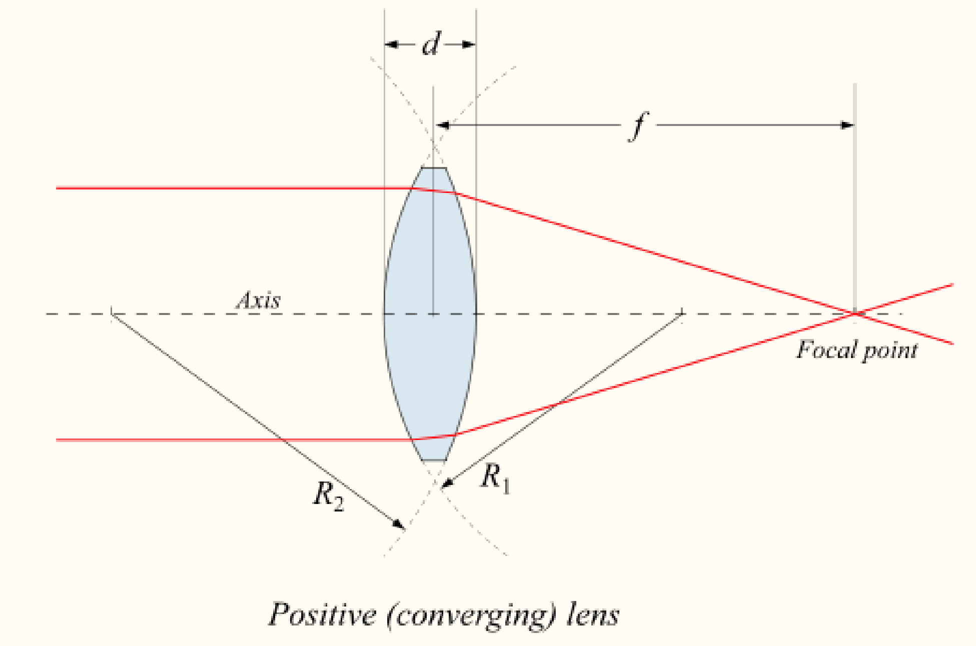
Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- **Basic optics**
- Camera models
- Camera calibration

Optical lenses

The optical element of an image acquisition device is one of the most important elements of these devices. They can be made of complex groups of lenses, in particular when variable zooming is desirable. Optical component study goes far beyond the aim of this course.

Since most simple image acquisition systems, however, use a single converging lens as its optical interface, we will look simply into this case.

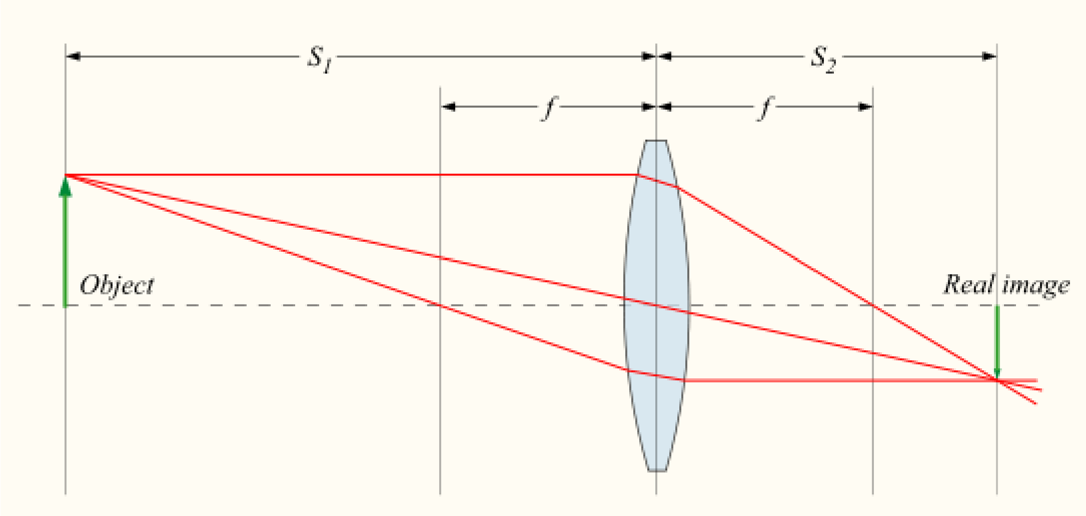


Formation of image

If the distances from the object to the lens and from the lens to the image are S1 and S2 respectively, for a lens of negligible thickness, in air, the distances are related by the thin lens formula which is an acceptable approximation to the full optical equation

$$\frac{1}{S_1}+\frac{1}{S_2}=\frac{1}{f}$$

If S1 >> S2 then S2 ~= f

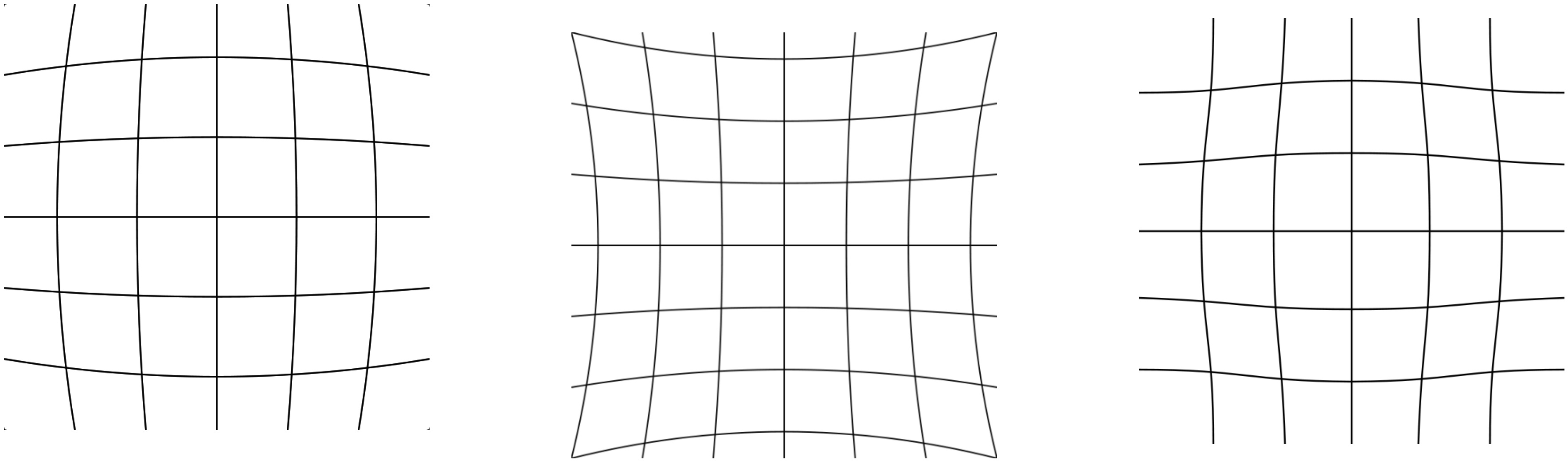


Spherical Aberration

Spherical Aberration is an image deformation resulting from the fact that most lenses have a spherical surface cut which is easier to make than its proper shape.

Spherical Aberration can show as a barrel form (left side), pincushion effect (center) and Mustache distortion (right).

The barrel distortion is by far the most common one and normally increases with diminishing focal distance.



Spherical Aberration

Spherical Aberration results in multiple focus points depending on the distance at which each ray of light enters the lens when referred to its major axis, and can be corrected by can be corrected using the simplified Brown's distortion model

$$x_u = (x_d - x_o)(1 + K_1 r^2 + K_2 r^4 + \ldots)$$
$$y_u = (y_d - y_o)(1 + K_1 r^2 + K_2 r^4 + \ldots)$$

where

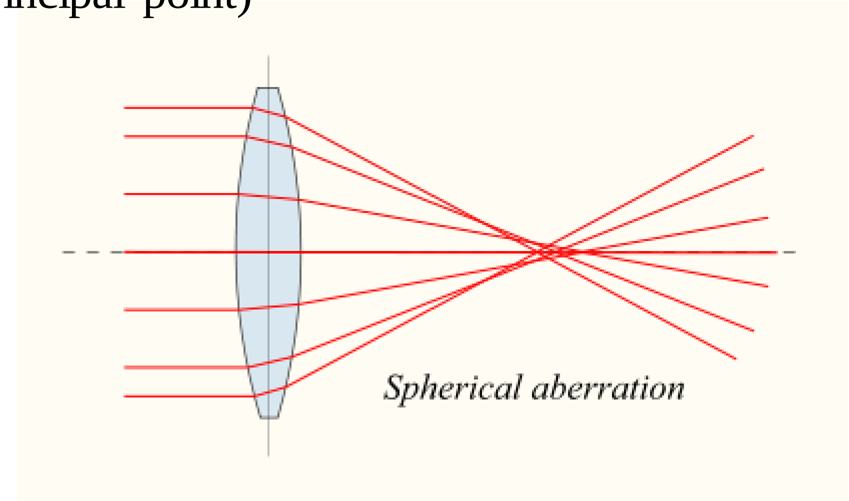
(x_d, y_d) - distorted image point as projected on image plane using specified lens

(x_u, y_u) - undistorted image point as projected by an ideal pin-hole camera

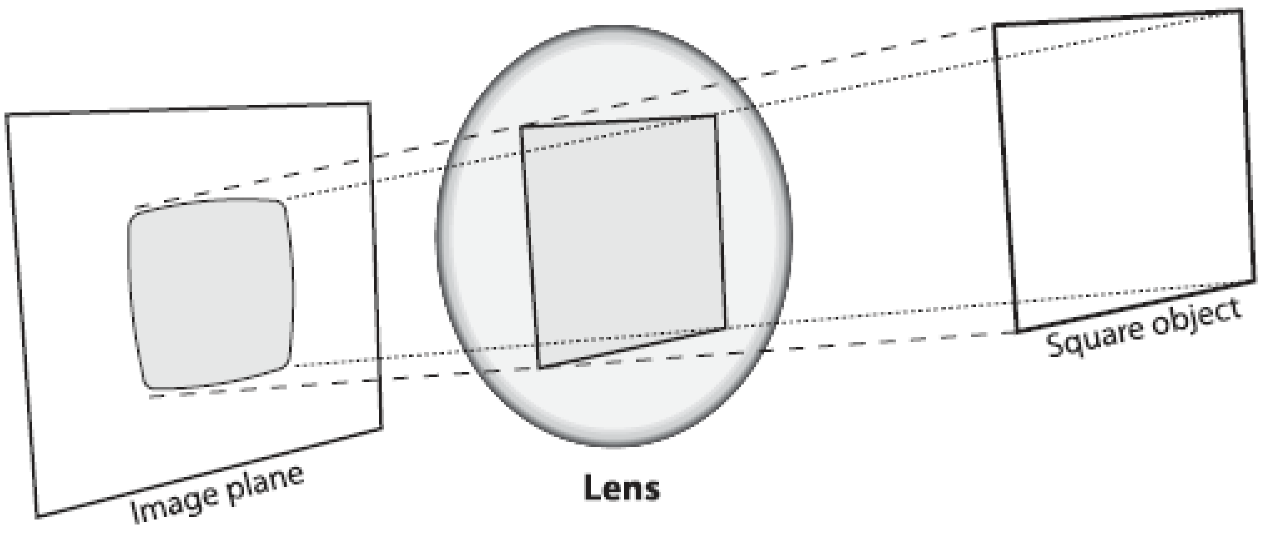
(x_o, y_o) - distortion center (assumed to be the principal point)

$K_n = n^{th}$ - radial distortion coefficient

$$r = \sqrt{(x_d - x_o)^2 + (y_d - y_o)^2}$$



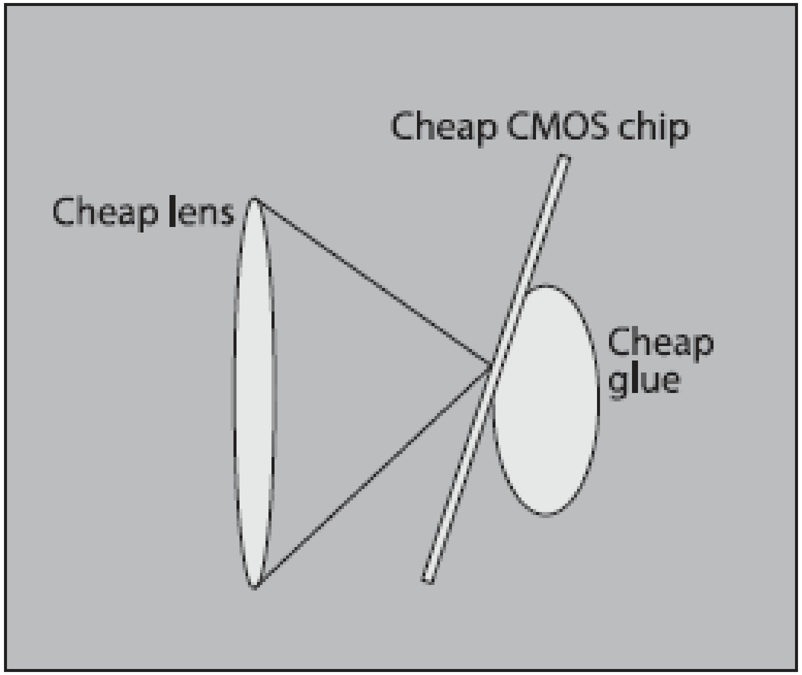
Lens distortion – Radial Distortion



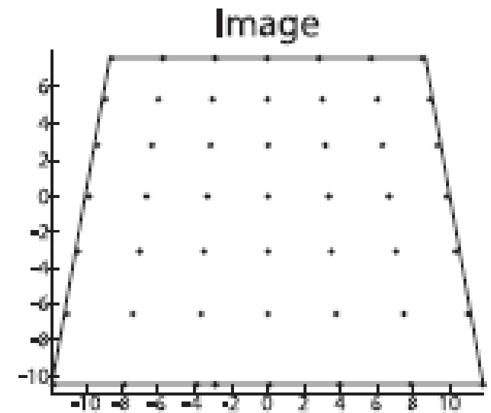
- $x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$
- $y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$

Lens distortion – Tangential Distortion

– Tangential



Cheap camera



– $x_{corrected} = x + [2p_1 y + p_2(r^2 + 2x^2)]$

– $y_{corrected} = y + [p1(r^2 + 2y^2) + 2p_2 x]$

Translation and rotation matrices

- Necessary also to find out the Pose (rotation and translation relative to a given coordinate system) of cameras
  - Rotation – vector with 3 rotation angles (r_x, r_y, r_z) that might be composed in a 3x3 rotation matrix.
  - Translation – vector with 3 values (t_x, t_y, t_z)

Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Parameters of the general model

Referring to the pinhole camera model, a camera matrix can therefore be used to denote a general projective mapping 3D world coordinates with 2D homogeneous coordinates of pixels in images.

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

Where A represents the camera intrinsic parameters, R and T the rotation and translation matrix $\boldsymbol{x_n}$ each of the coordinates in 3D space and ***u*** and ***v*** the indexes of each pixel on the frame buffer.

$$\mathrm{A} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

OpenCV Camera model

- OpenCV Camera model:
  - 4 intrinsic parameters:
    - Focal distance: f_x, f_y
    - Optical centre: c_x, c_y
  - 5 distortion parameters
    - Lens distortion: k_1, k_2, k_3, p_1, p_2
  - 6 extrinsic parameters:
    - Rotation: r_x, r_y, r_z
    - Translation: t_x, t_y, t_z

Total: **15 parameters**

- Other models: Tsai, Heikkila, Zhang

Sumary

- Basic definitions
- The image processing pipeline
- Image parameters
- Camera parameters
- Basic optics
- Camera models
- Camera calibration

Single camera calibration

- Given 3D-2D correspondences, possible to define a system of equation that might be solved numerically.

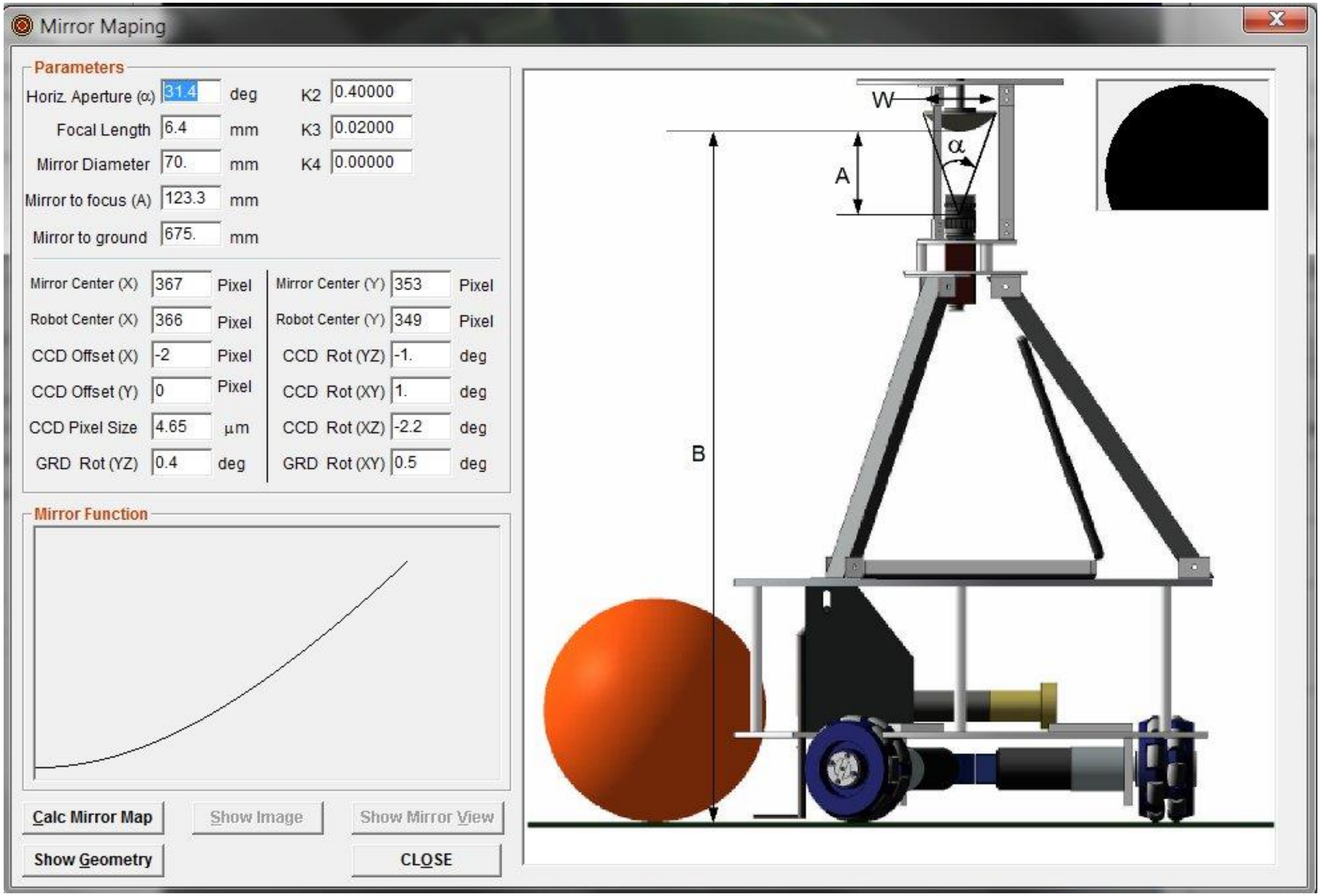
- RANSAC "RANdom SAmple Consensus” might be used to remove outliers

Camera calibration OpenCV

- cvFindChessboardCorners: internal corners of the chessboard
  - Optional: cvFindCornerSubPix: Refines the corner locations
- cvCalibrateCamera2: find intrinsic and extrinsic parameters from several views of chessboard.
- solvePnP: finds an object pose from 3D-2D point correspondences.

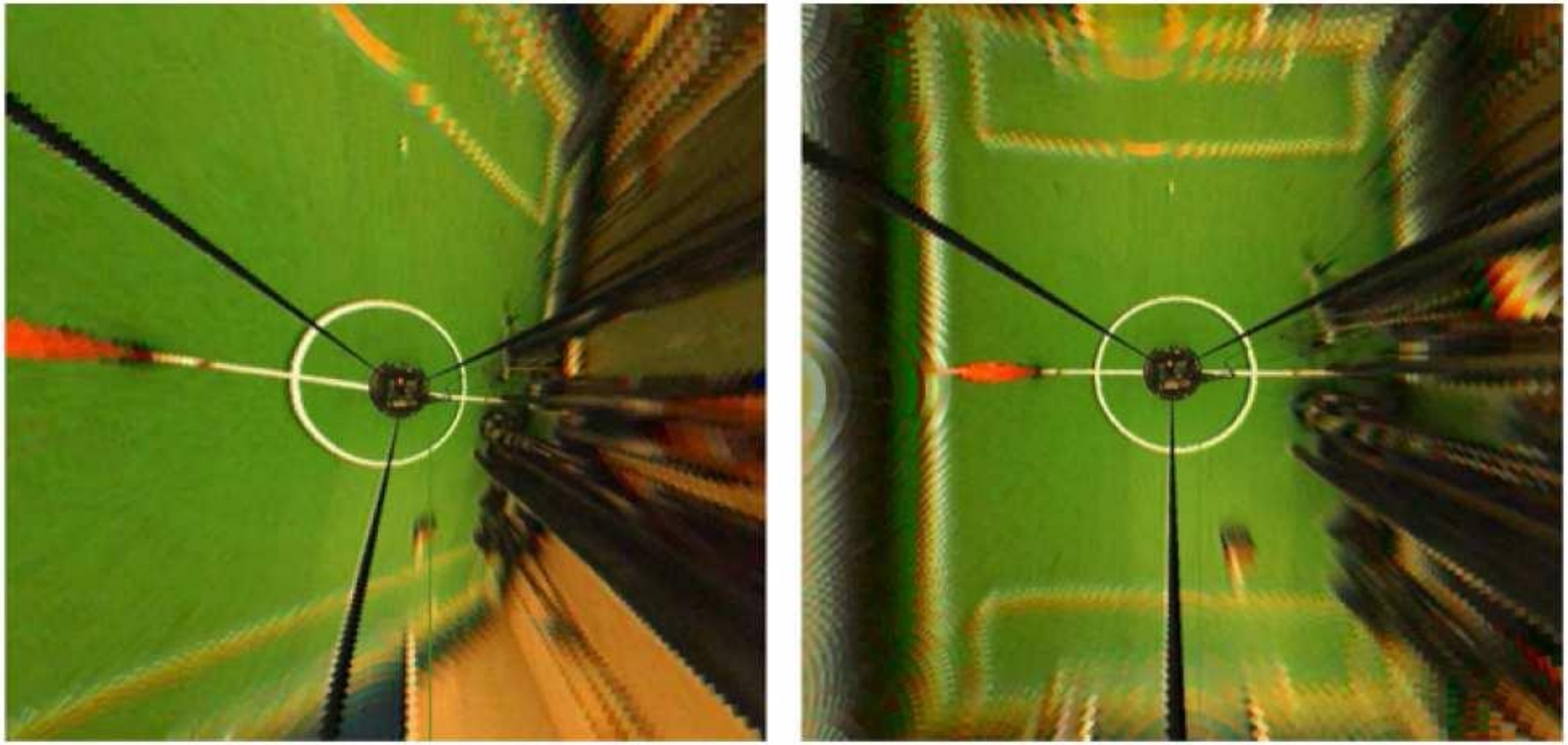
Vision System – example

- A catadioptric system is a vision system based is based on one camera and one mirror system



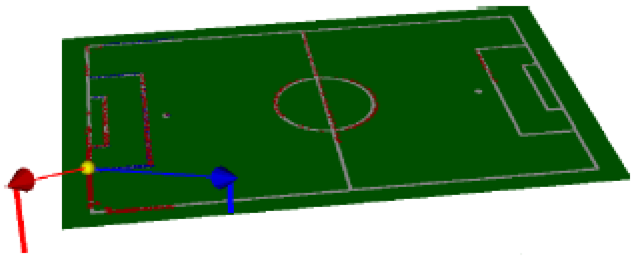
Vision System –example

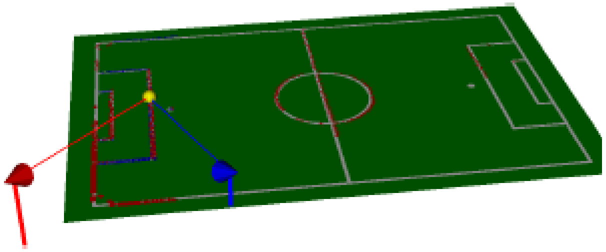
Such a system suffers from mechanical misalignments originating from several possible different reasons. By using the camera model, the catadioptric model and estimating the geometry parameters that introduce the distortion, one can correct the original image to obtain a fairly accurate distance map at the ground level

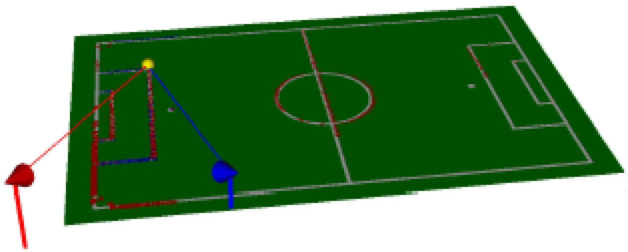


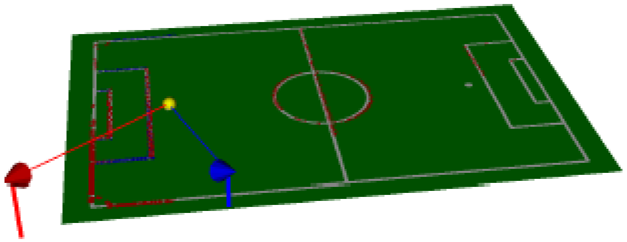
Vision System – example

Ground truth for ball positions







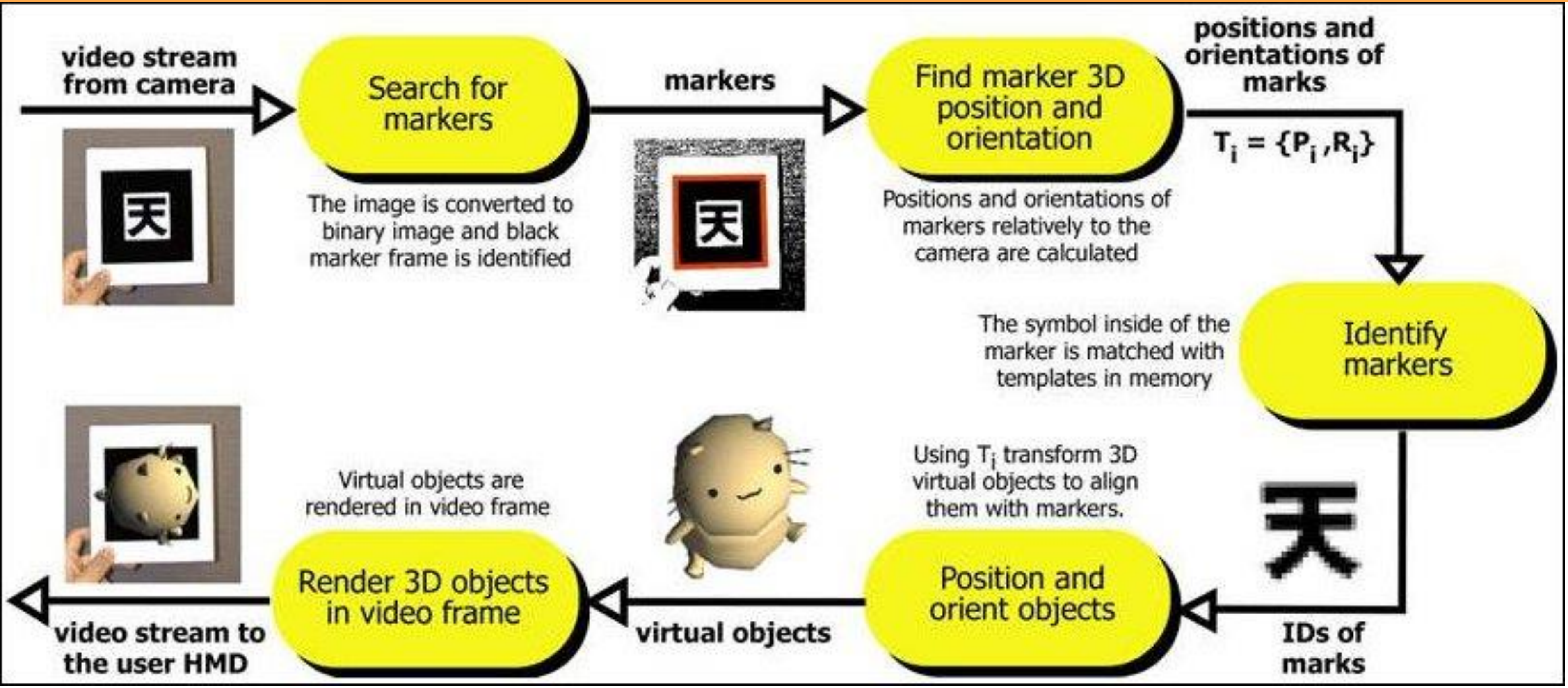


OPENCV CAMERA CALIBRATION DEMO

ARToolkit

- C/C++ library
- Estimates position/orientation of a camera
- Marker detection
- Camera calibration procedure
- Allow Real Time AR application
- 3D models reading
- Open-source

ARToolkit principle



- captures video of the real world
- searches through each video frame for any square shapes.
- If a square is found, mathematics to calculate the position of the camera relative to the square.
- graphics model is drawn from that same position.
- model is drawn on top of the video of the real world and so appears stuck on the square marker.
- final output is shown back in the display, user looks through the display and see graphics overlaid on the real world.

ARToolkit – Camera Calibration

- Default camera properties are contained in the camera parameter file camera_para.dat, that is read in each time an application is started.
- should be sufficient for a wide range of different cameras.
- However using a relatively simple camera calibration technique it is possible to generate a separate parameter file for the specific cameras that are being used (http://www.hitl.washington.edu/artoolkit/documentation/usercalibration.htm)
- if camera parameters are known then the video image can be warped to remove camera distortions.

Some references

- Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, Cambridge, MA, 2008.
- Olivier Faugeras Three-dimensional computer vision: a geometric viewpoint. MIT Press Cambridge, MA, USA ©1993
- Szeliski, R. (2010). Computer Vision: Algorithms and Applications, Springer
- Quang-Tuan Luong. "Learning Epipolar Geometry". Retrieved 2007-03-04.
- Richard Hartley and Andrew Zisserman (2003). Multiple View Geometry in Computer Vision. Cambridge University Press. pp. 155–157. ISBN 0-521-54051-8.
- Z. Zhang, "A flexible new technique for camera calibration"', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.11, pages 1330–1334, 2000.