Computer Vision basics: Digital Cameras Parameters

Institute of Electronics and Telematics Engineering of Aveiro
Dep. of Electronics, Telecomunications and Informatics
Universidade de Aveiro
Portugal

Outline

- Broad definition of computer vision & artificial vision areas
- Some basic definitions
- The typical image processing pipeline
- Image parameters (intrinsic and adjustable)
- Basic optics
- Camera parameters and perspective views
- Examples of simple image processing tasks

S Complementary slide

Broad definition

- Computer vision encompasses a wide field that includes methods for acquiring, processing, analyzing, and understanding images.
- Computer vision applications are always increasing:
  - surveillance;
  - machine inspection;
  - medicine;
  - **robotics**;
  - entertainment;
  - media.
- One of the goals: make computer vision converge towards human vision (or make it even better).

Artificial vision areas

- Artificial Vision
  - Wide research field that includes all sciences and techniques that allow the study and application of all activities related to the use and interpretation of an image.
- Computer Vision
  - Techniques for image acquisition, extraction, characterization and interpretation of the information gathered from images of the 3D world.
  - **Machine Vision**
    - Computer Vision for automation and robotic applications.
- Image Processing
  - Signal processing where the input signal is an image (a 2 or 3 dimensional signal) and the output can be a transformation of this image, or a set of characteristics associated to the image.

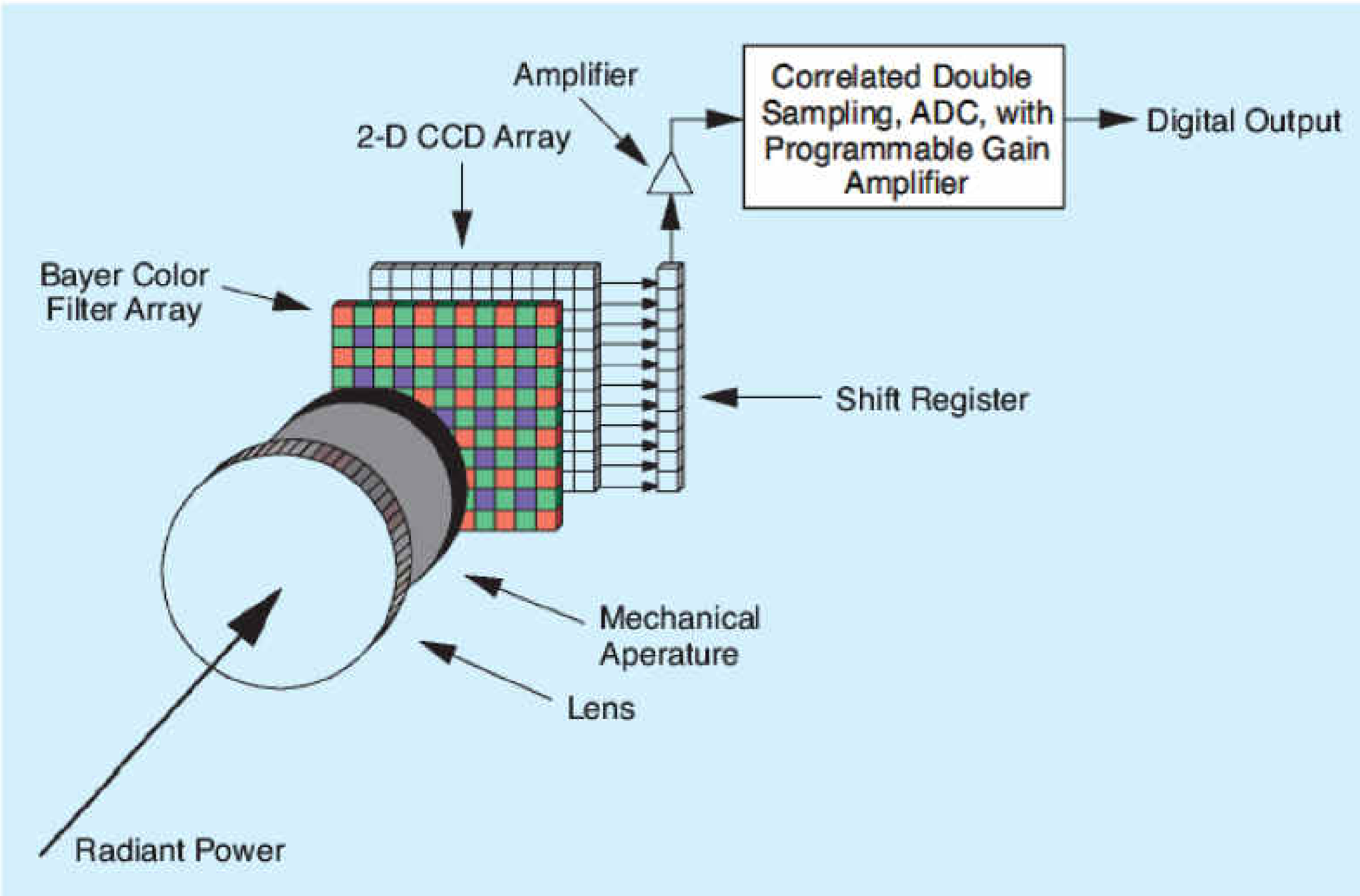
Computer vision

- Broad definition of computer vision & artificial vision areas
- **Some basic definitions**
- The typical image processing pipeline
- Image parameters (intrinsic and adjustable)
- Basic optics
- Camera parameters and perspective views
- Examples of simple image processing tasks

Digital Cameras

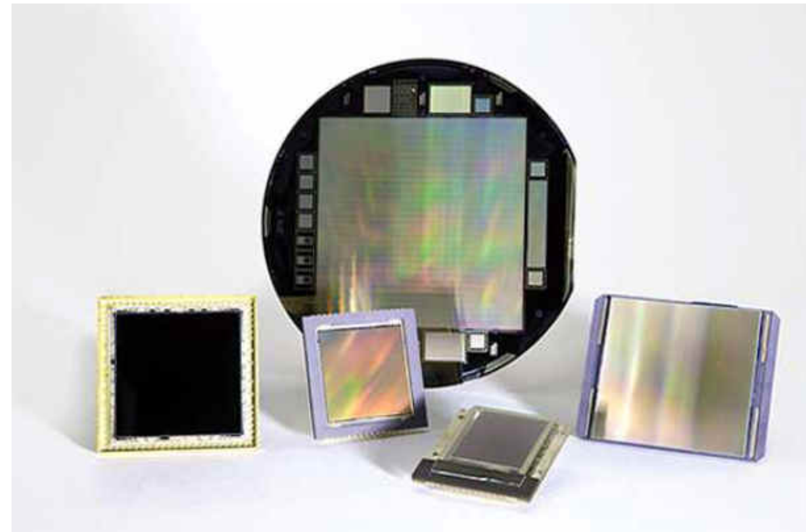
Image acquisition using a digital camera (general case nowadays):

(IEEE SP Magazine, Jan 2005)



Digital Cameras (sensors)

- Some considerations: speed, resolution, cost, signal/noise ratio, . . .
- **CCD** - Charge Coupled Device
  - Higher dynamic range
  - High uniformity
  - Lower dark noise.
  - More expensive and complex.



- **CMOS** - Complementary Metal Oxide Semiconductor
  - Lower voltage, lower power
  - Higher speed
  - Lower system complexity; pre-processing capabilities.
  - Very flexible in terms of in camera complementary hardware

Digital Cameras variants

- Several interfaces (Firewire, GigE, CameraLink, USB, . . .).
- Scientific usage (high resolution, long exposure time, . . .).
- High speed (ex. >>1000 fps).
- Linear (ex. 10000 lines/second).
- 3D
- Infrared (ex. 8 to 14 µm).
- Multispectral
- High dynamic range
  (ex. using a prism and two sensors).

Basic 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/m^2.

CGS – Computer Graphic Systems

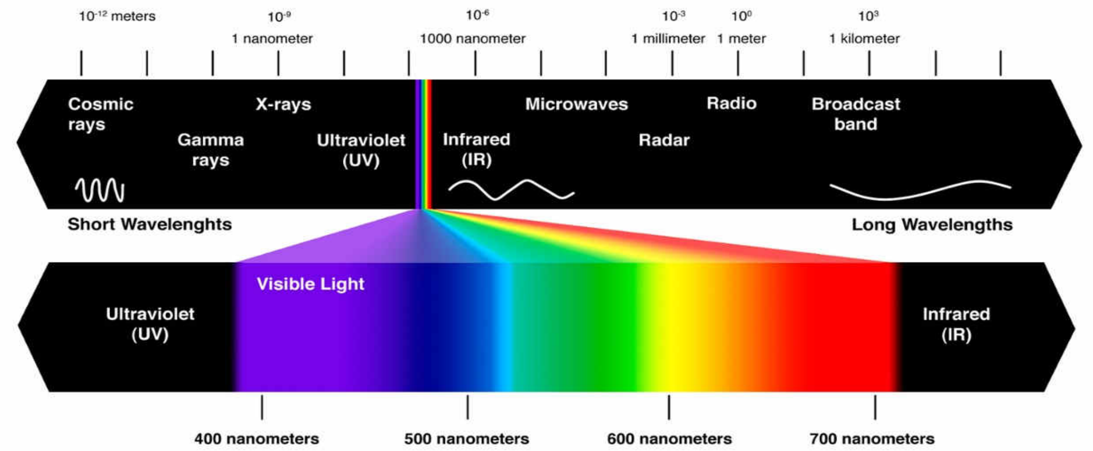
Chrominance

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

A black and white image (gray image) has a balanced distribution of energy among 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.



RGB & YUV

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

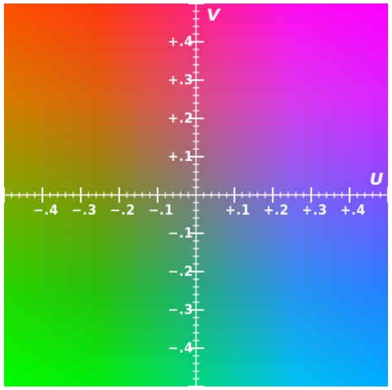
where

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

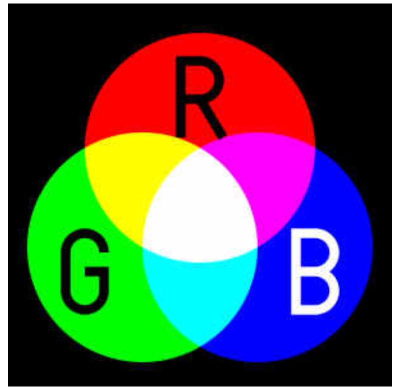
This values originally derivates from the general model of the human visual system and had a significant impact on the ability to develop a television color system compatible with the previous B&W television systems.

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

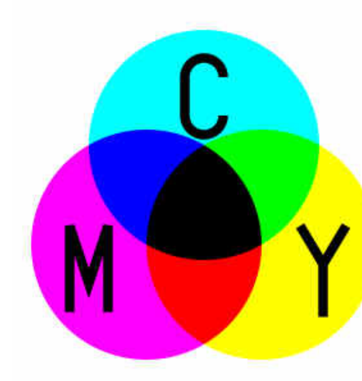
Examples of color spaces



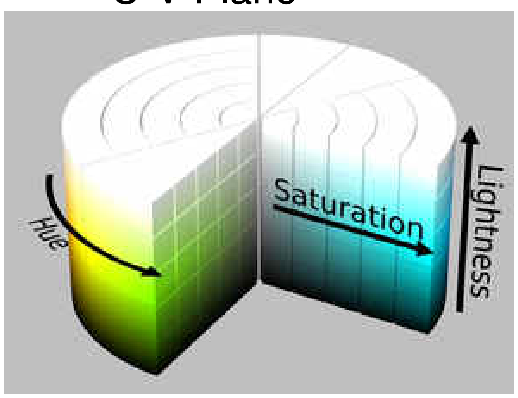
U-V Plane

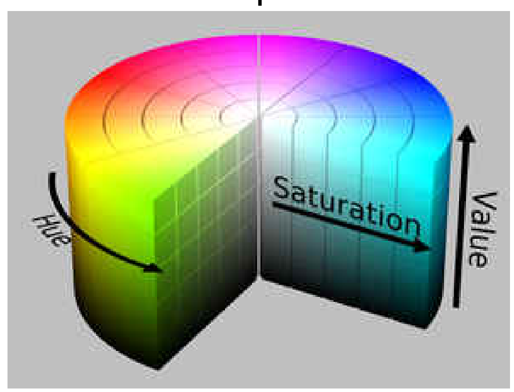


RGB Space



CYM(K) Space

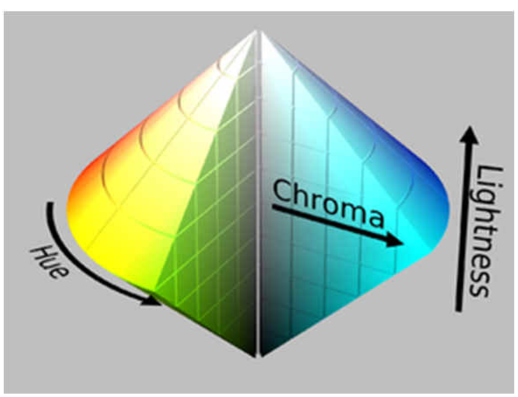


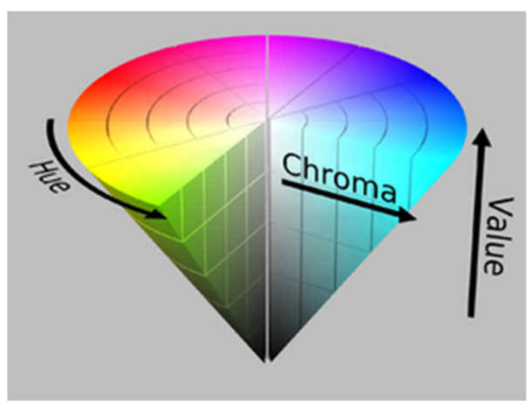


HSL and

HSV

Spaces





HCL and

HCV

Spaces

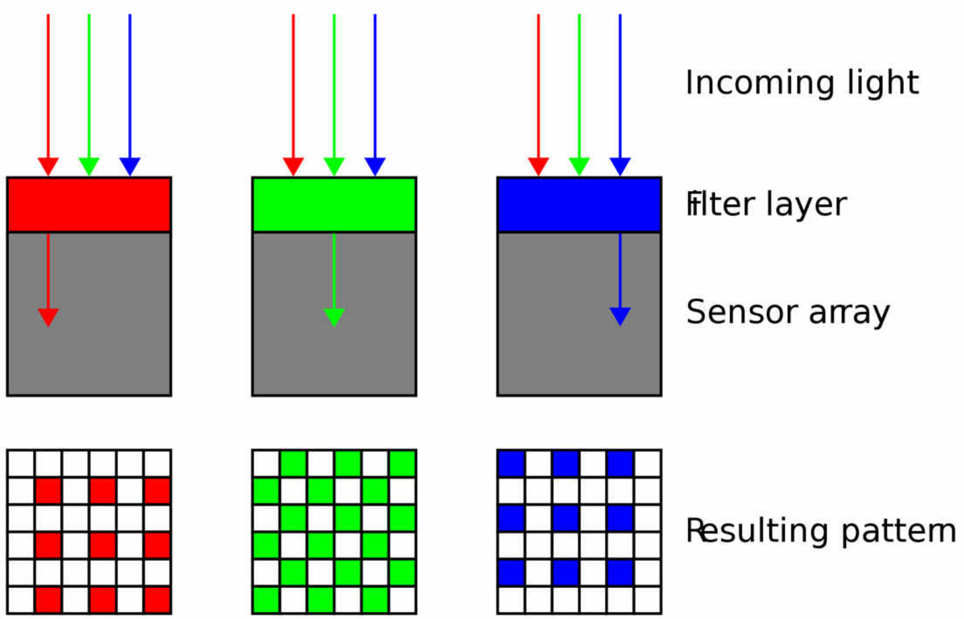
Image Sources: Wikipedia

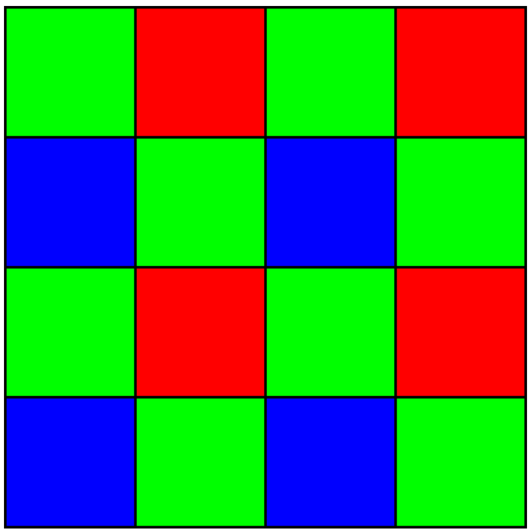
Image formation

Most image sensors use either a CCD or CMOS technology that is able to “grab” light energy through a conversion process from photons to electrons which are then trapped either in an array of potential wells or an array of capacitors.

Only luminance is captured in this way.

To obtain a color image a set of filters must be used, from which the Bayer pattern configuration is the most common.





Computer vision

- Broad definition of computer vision & artificial vision areas
- Some basic definitions
- **The typical image processing pipeline**
  - Image parameters (intrinsic and adjustable)
- Basic optics
- Camera parameters and perspective views
- Examples of simple image processing tasks

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.

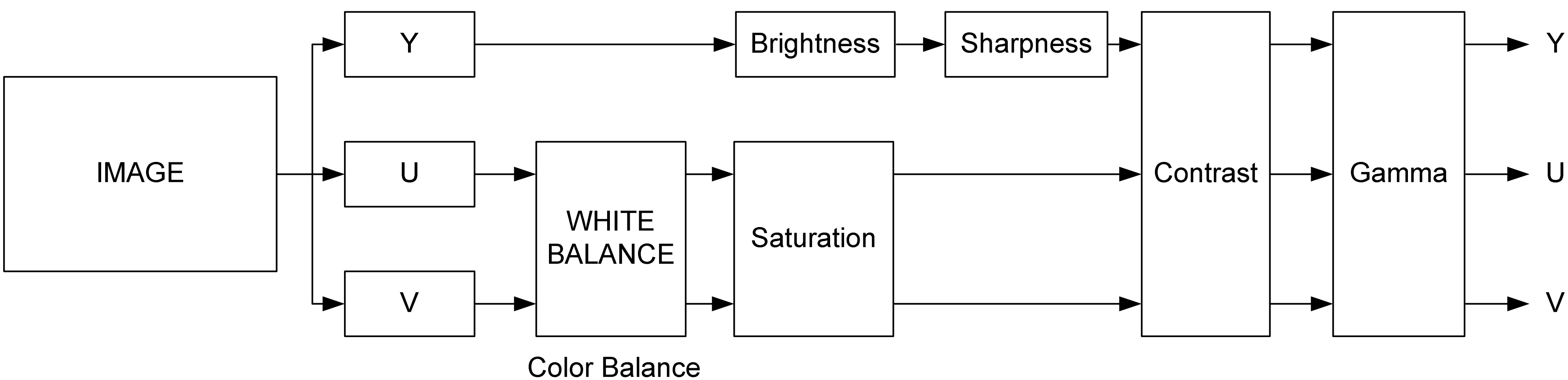
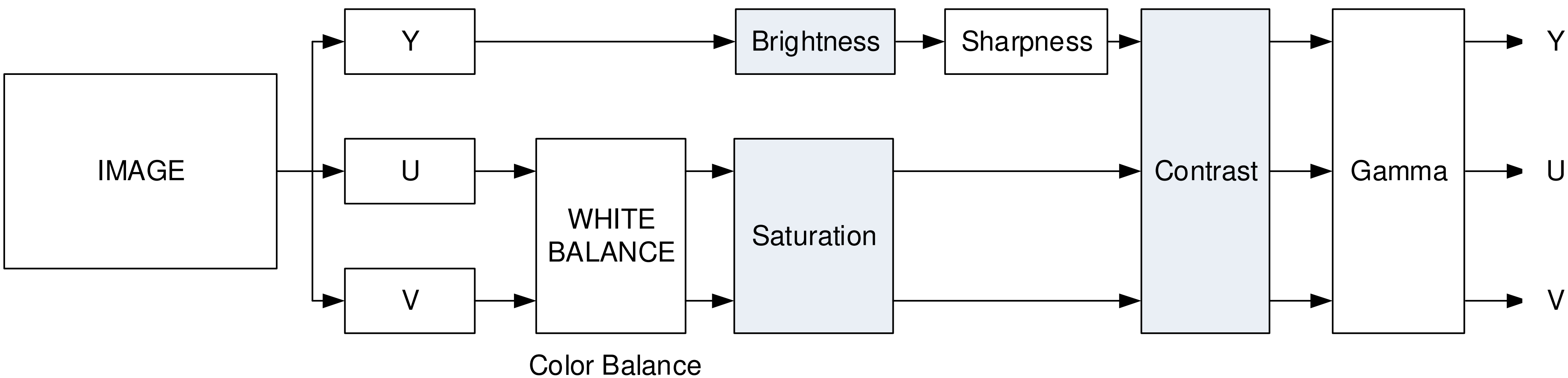


Image processing pipeline

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.

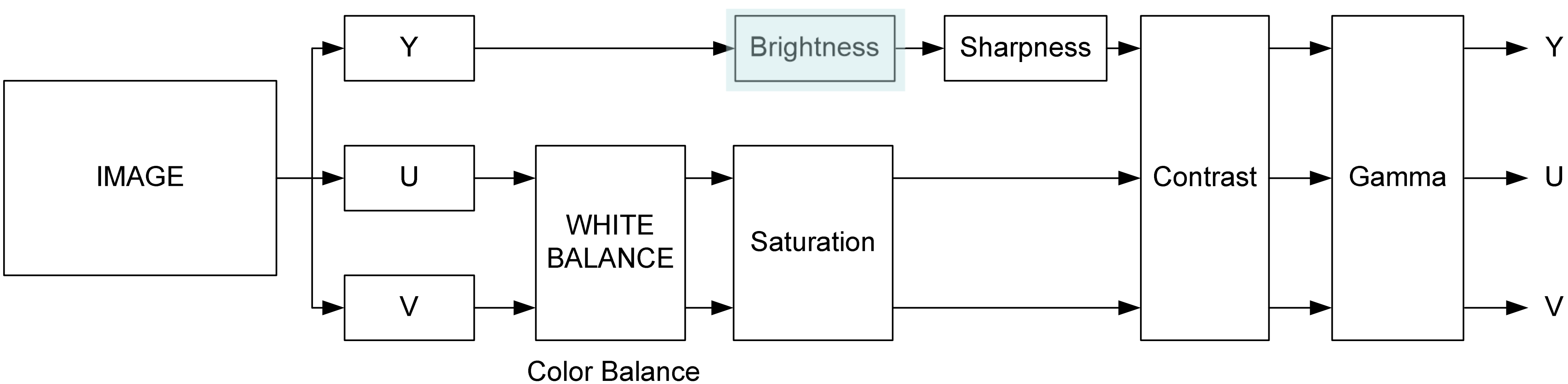


Brightness (as an intrinsic image characteristic)

Brightness 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).

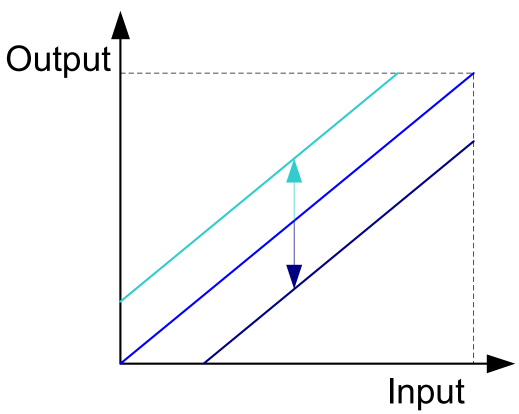
If the brightness it too high, overexposure may occur which will white saturate part or the totality of the image.

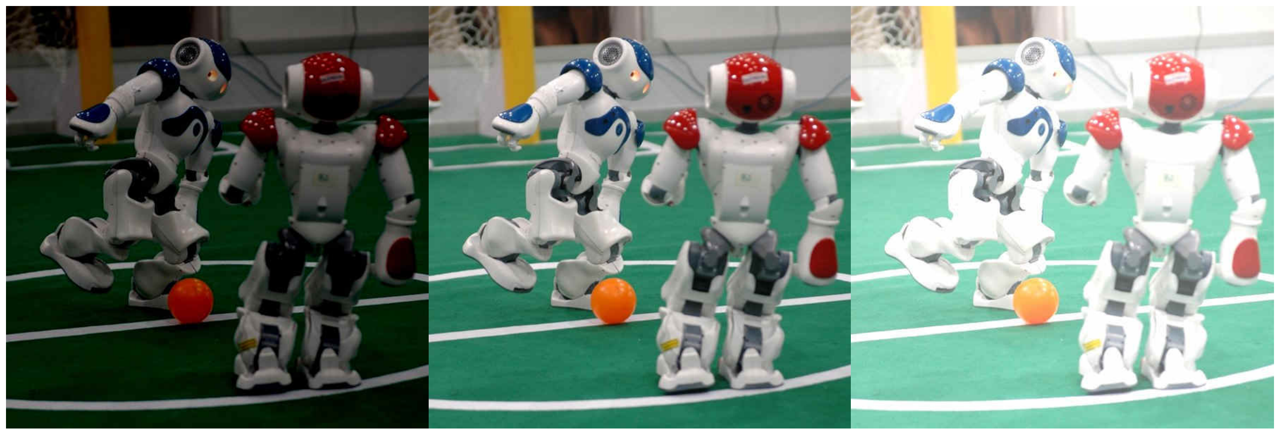


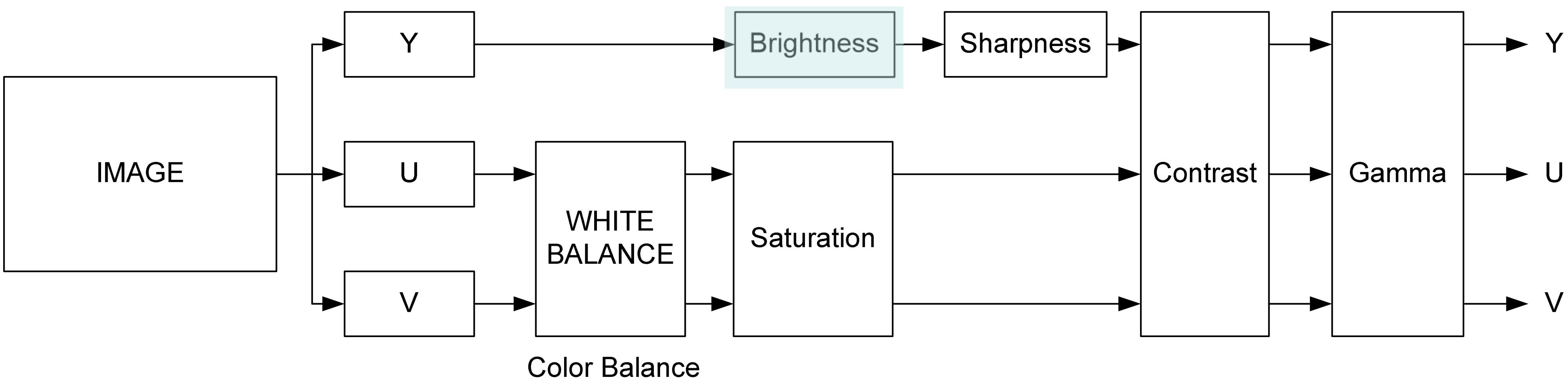
Brightness

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.







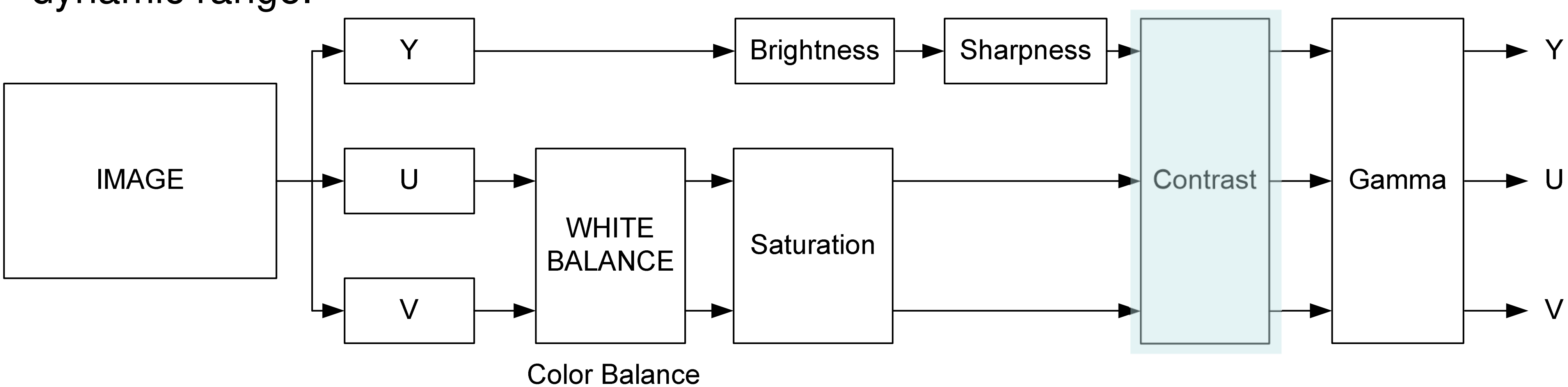
Contrast

Contrast (as an intrinsic image characteristic)

There is not a unique definition of contrast. One 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.

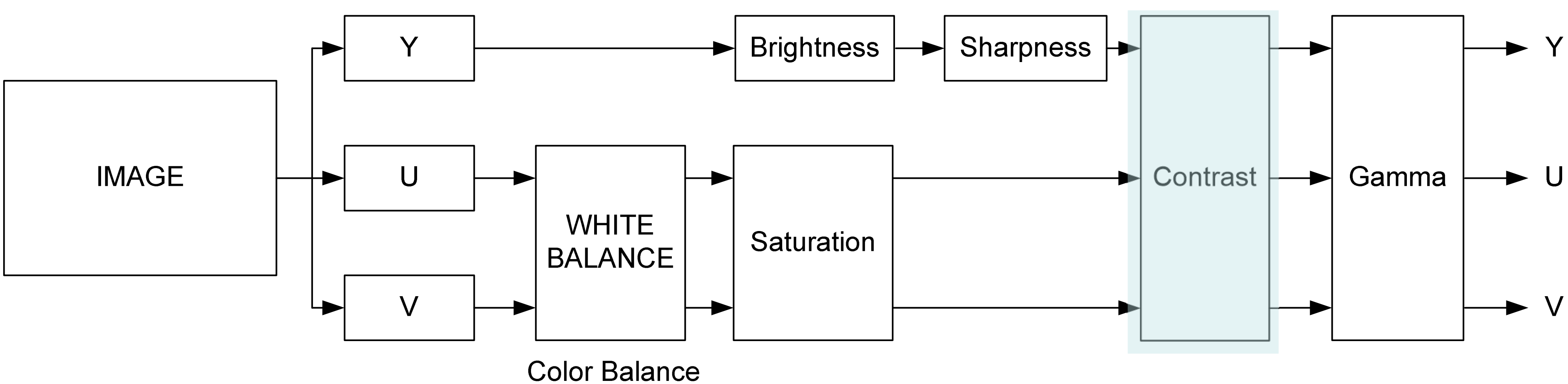


Contrast (as an intrinsic image characteristic)

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

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

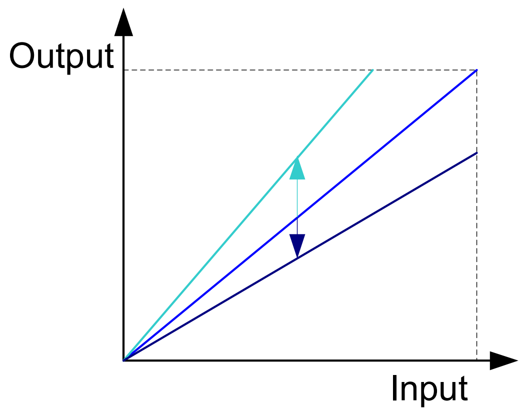
The human eye contrast sensitivity function is a typical band-pass filter with a maximum at around 4 cycles per degree with sensitivity reducing to both sides off that maximum. This means that the human visual system can detect smaller contrast differences at 4 cycles per degree than at any other spatial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

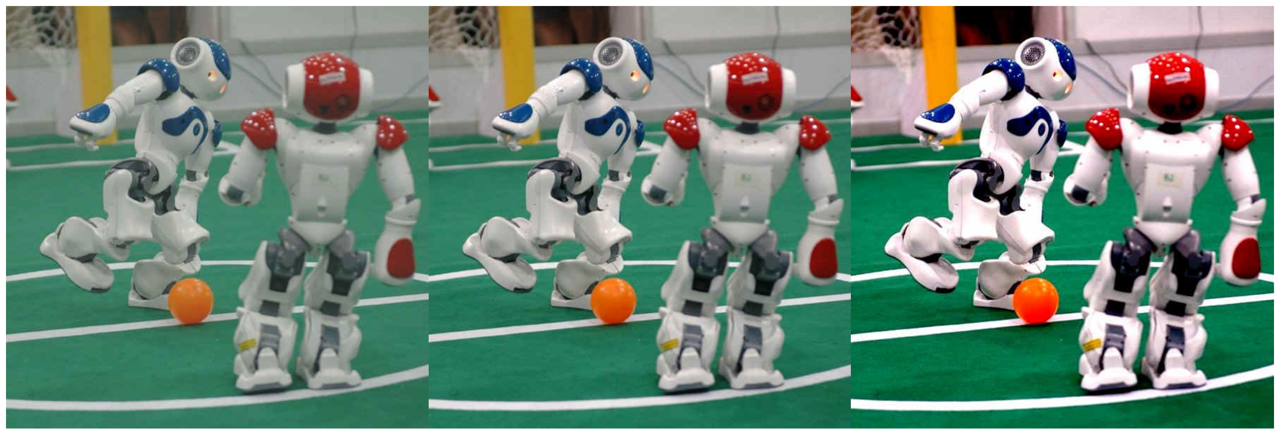


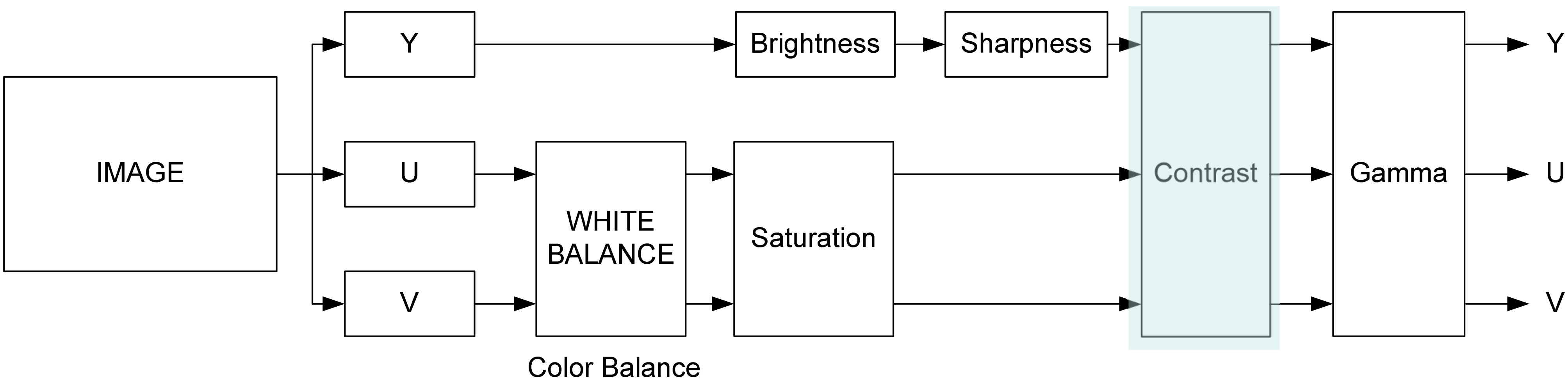
Contrast

Contrast (as a controllable parameter)

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



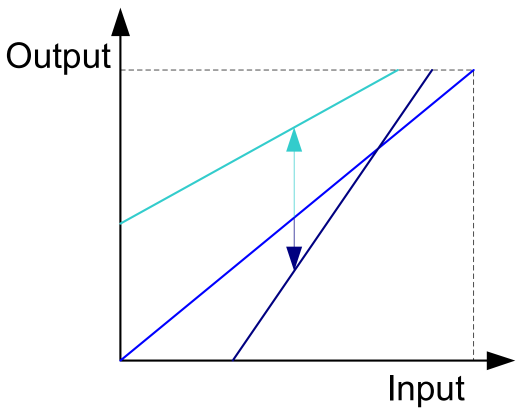


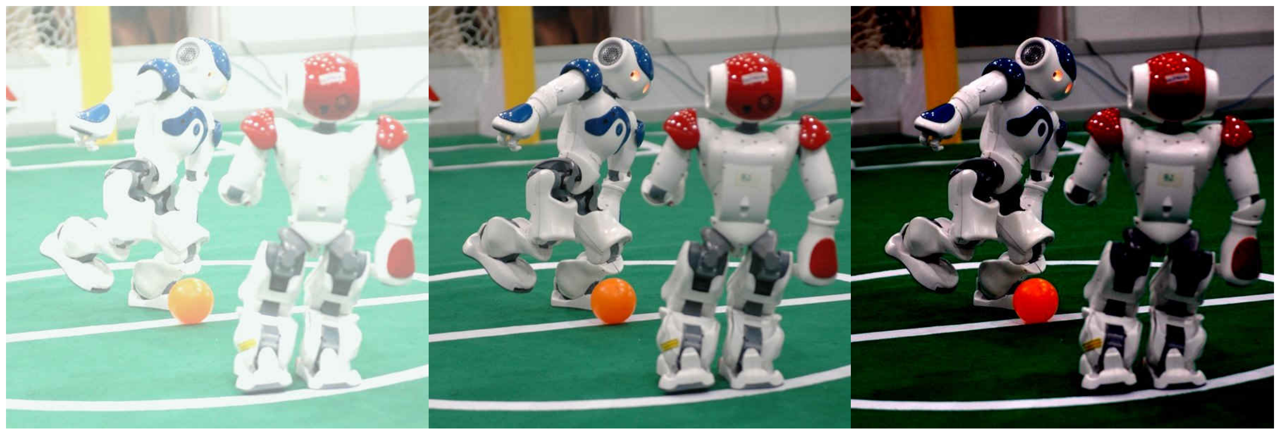


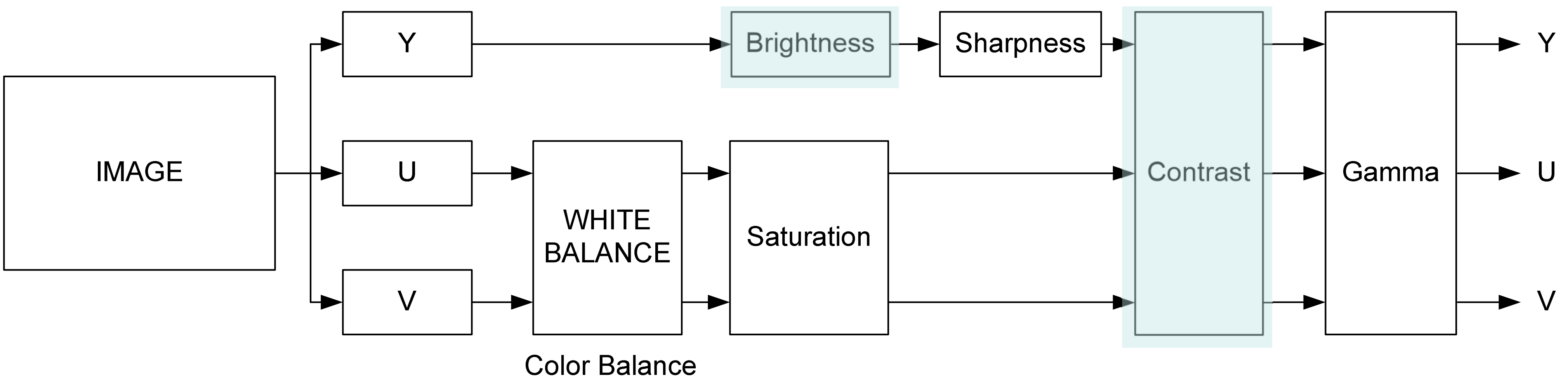
Contrast + Brightness

Contrast + Brightness(as controllable parameters)

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





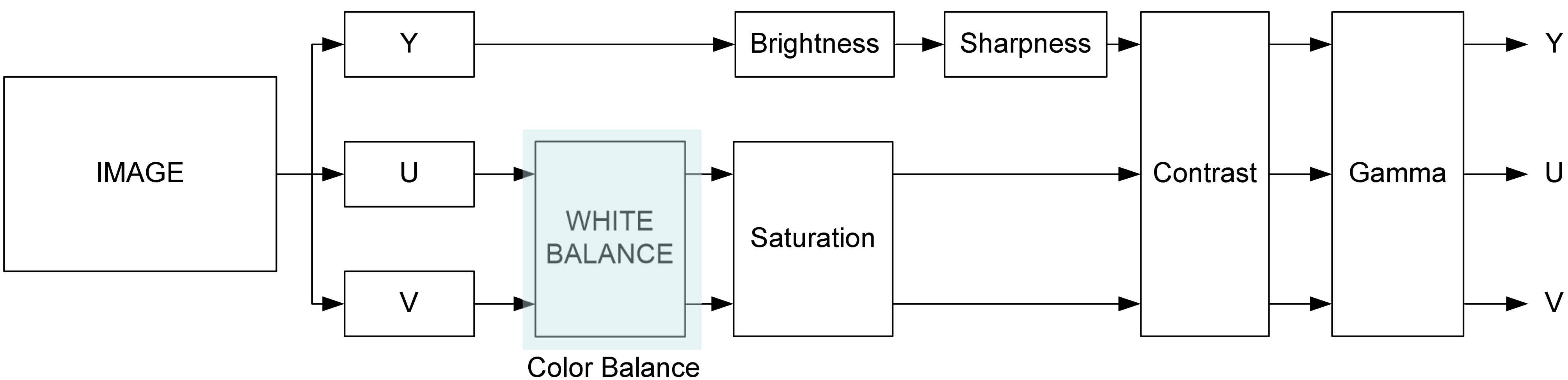


White Balance (as controllable parameters)

White balance is the global adjustment of the gain function 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.

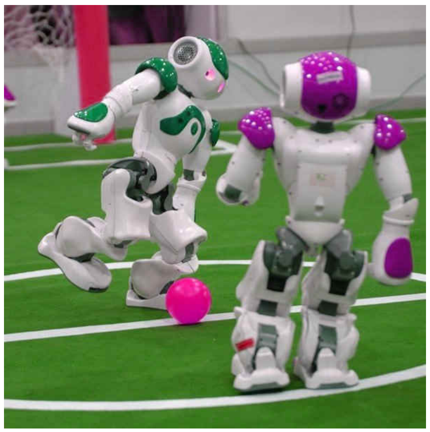
This balance is required because of different color spectrum energy distribution depending on the illumination source.

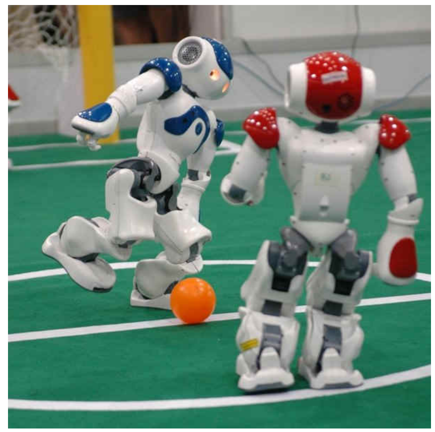


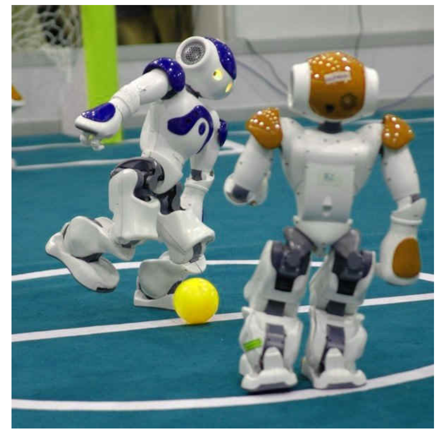
White Balance

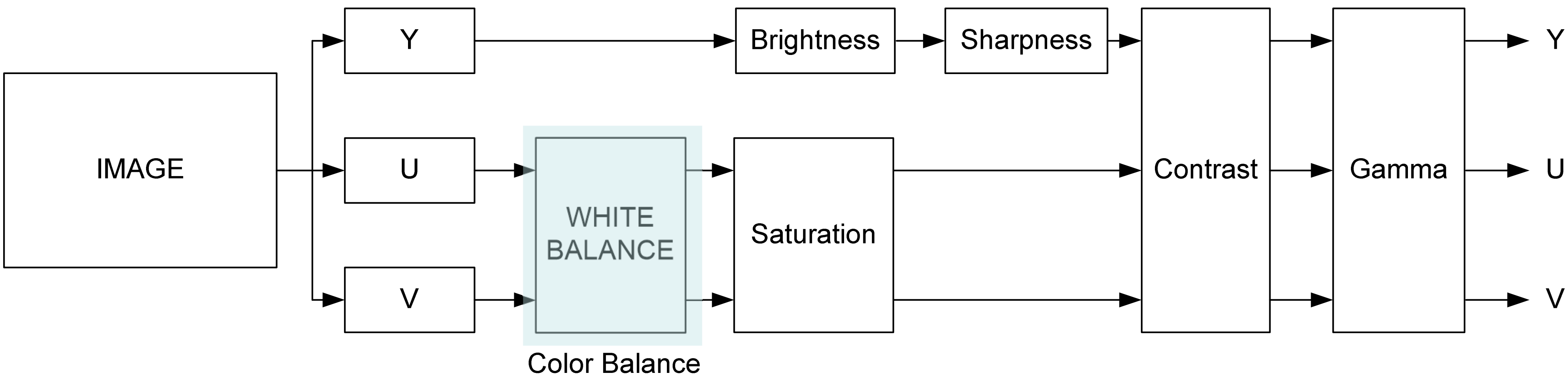
White Balance

Examples







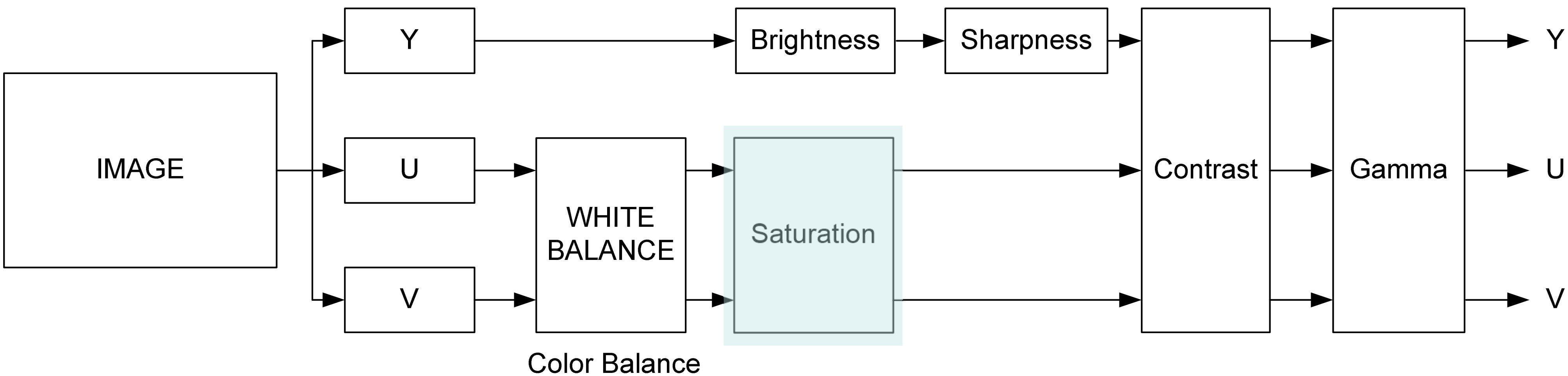


Saturation (as an intrinsic image characteristic)

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 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.

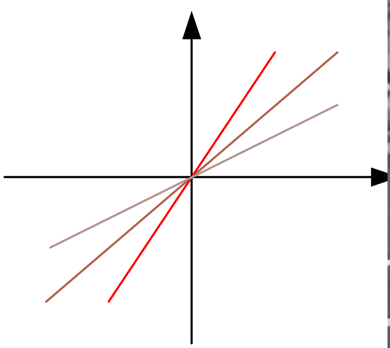
A non saturated image (B&W) has a spectrum distribution that matches the human eye spectrum sensibility. Saturation is sometimes also defined as the amount of white you have blended into a pure color.

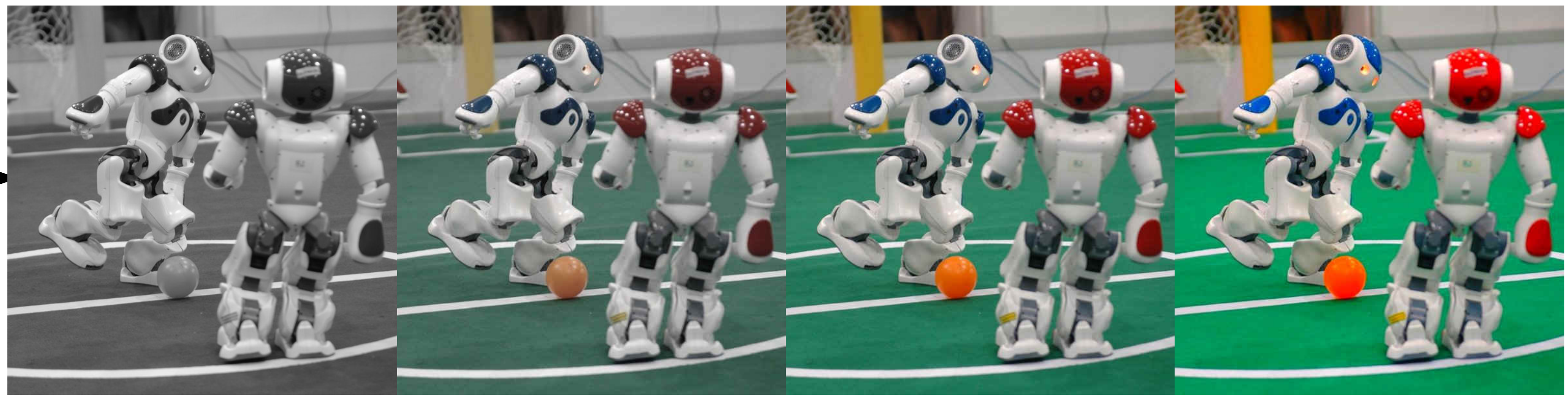


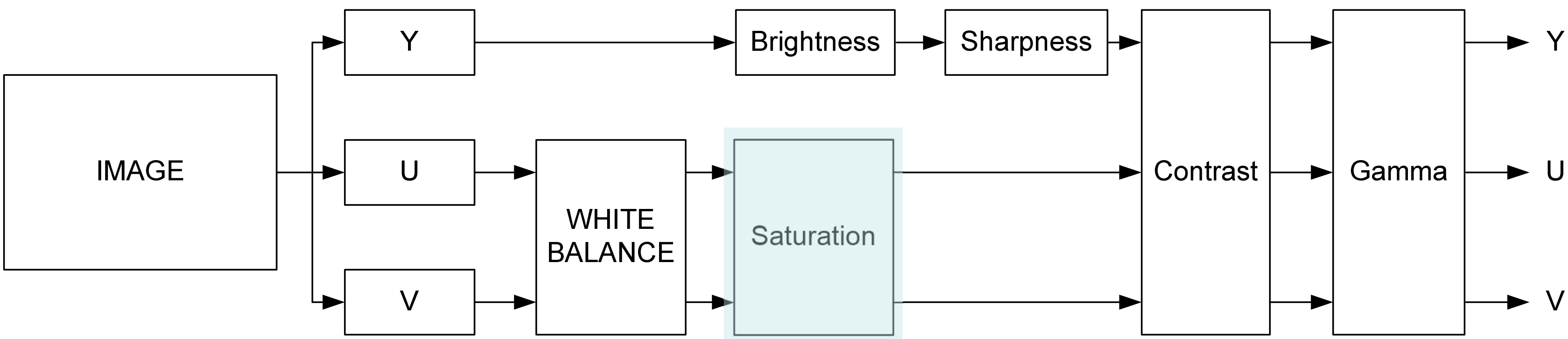
Saturation

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





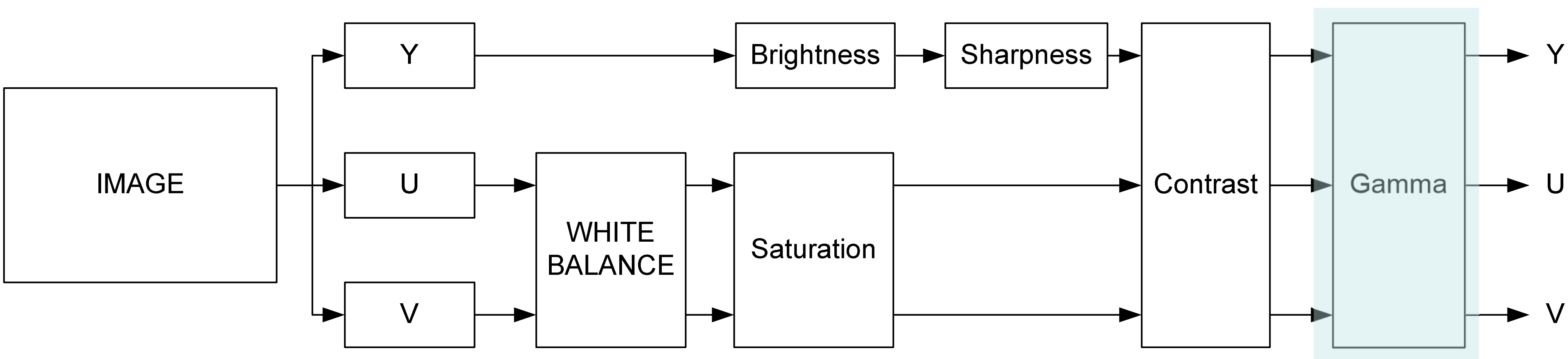
Gama

Gamma correction is the name of a nonlinear operation used to code and decode luminance or RGB 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.

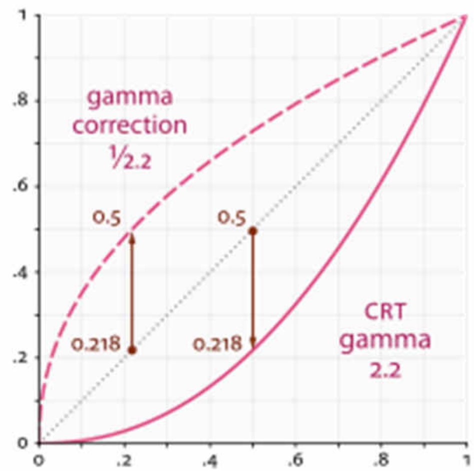


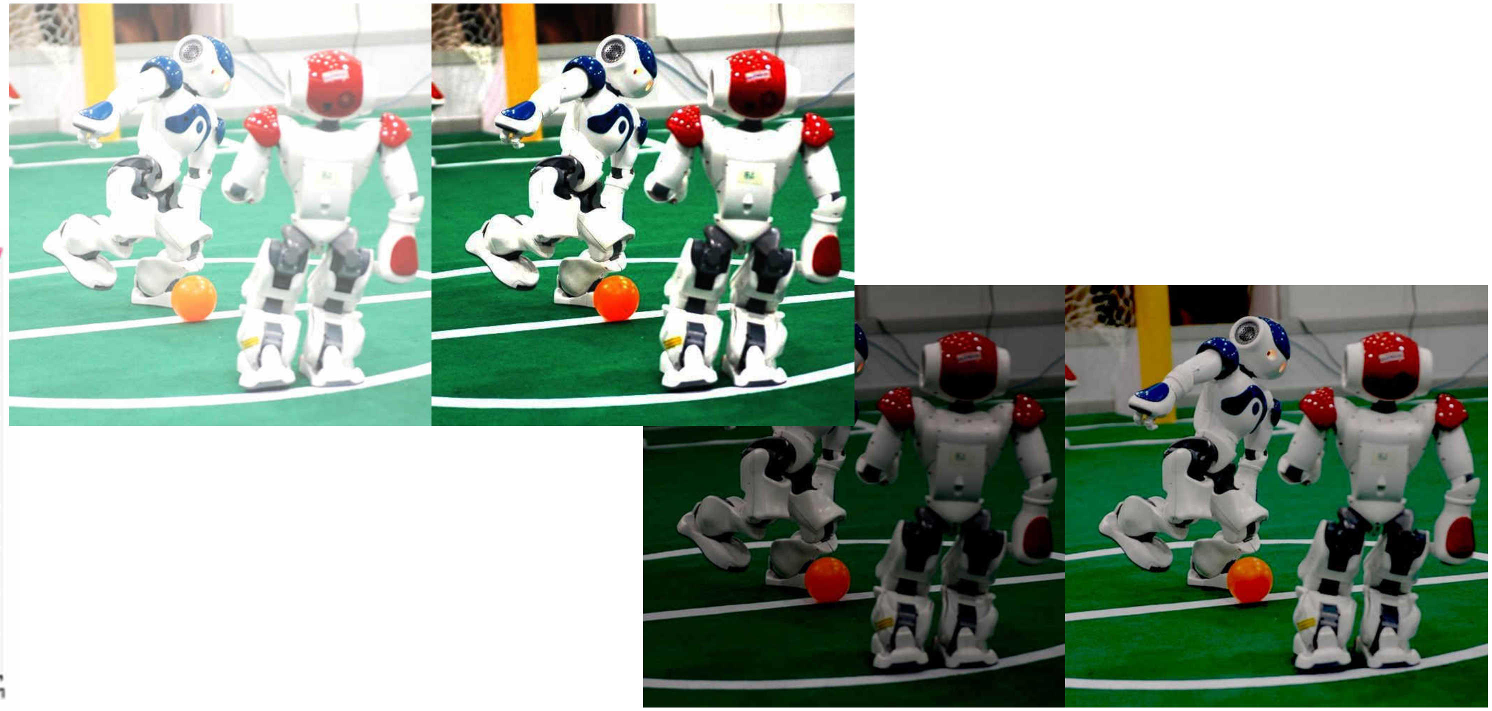
Color Balance

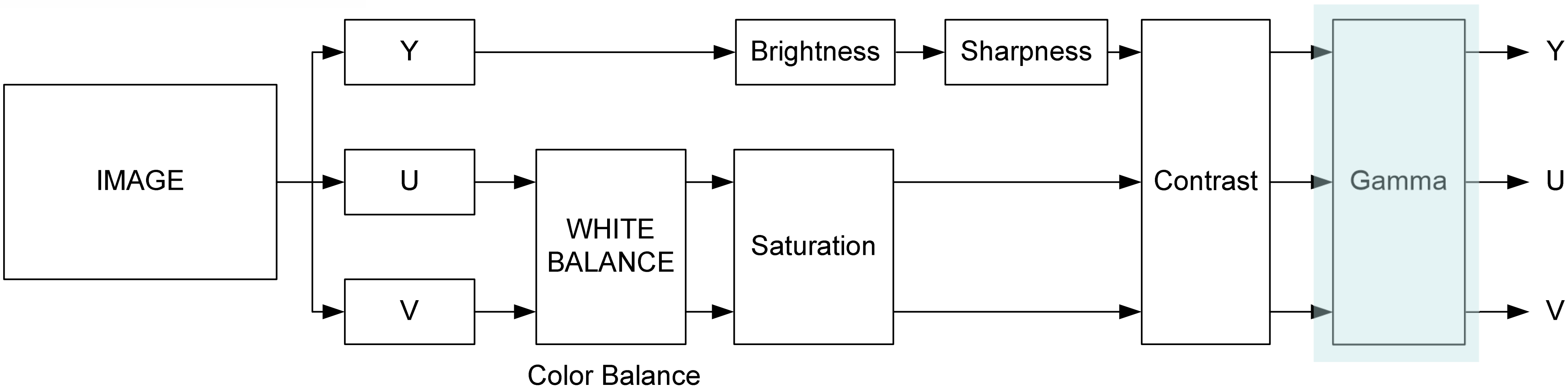
Gamma

Gamma

Examples







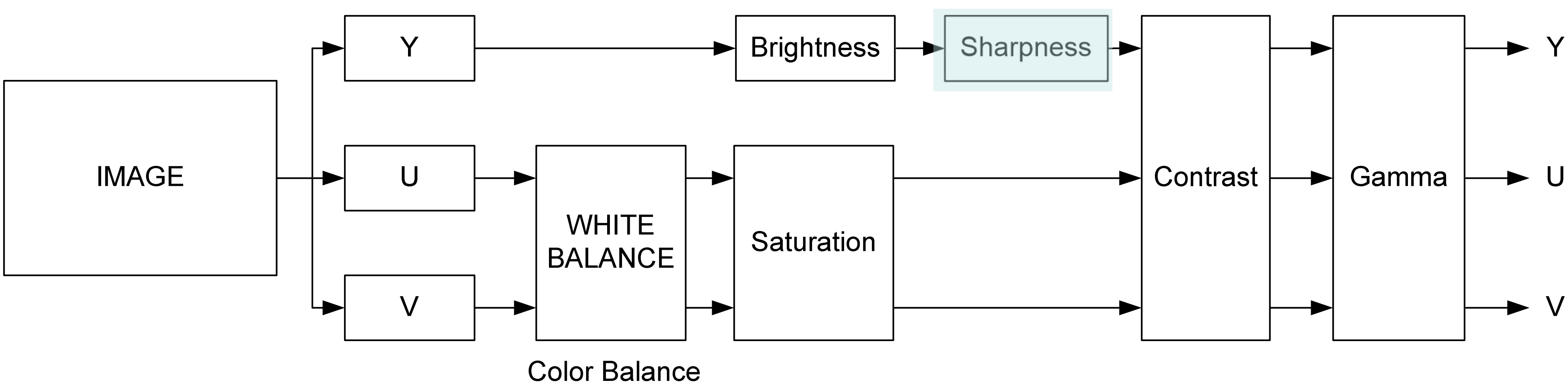
Sharpness

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.

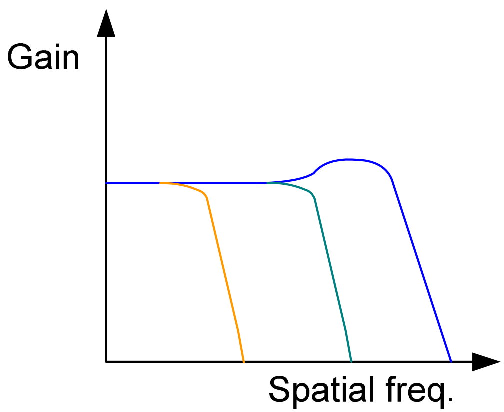
Sharpness basically allows the control of the cut-off frequency of a low pass spatial filter. This may be very useful if the image is afterward intended to be decimated, since it allows to prevent spatial aliases artifacts.

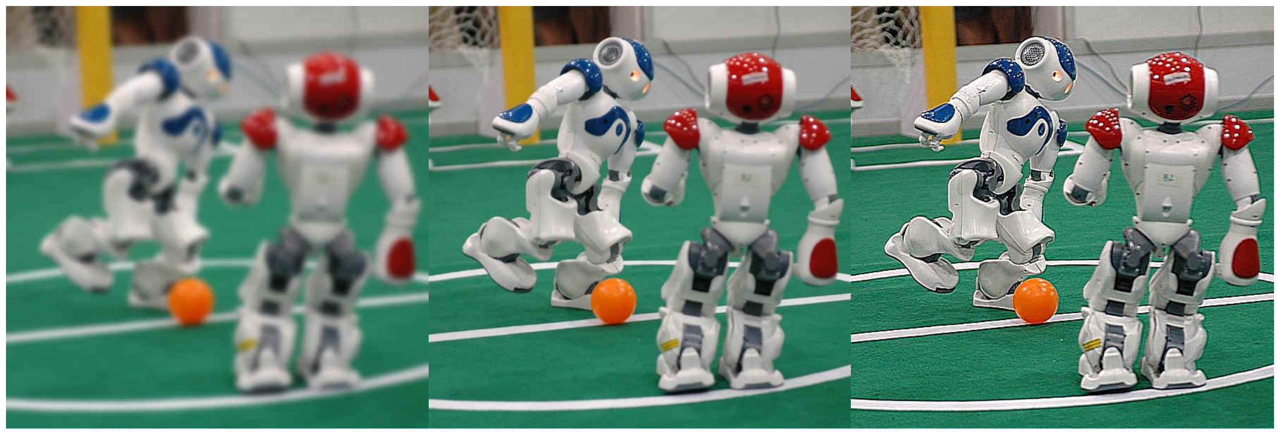


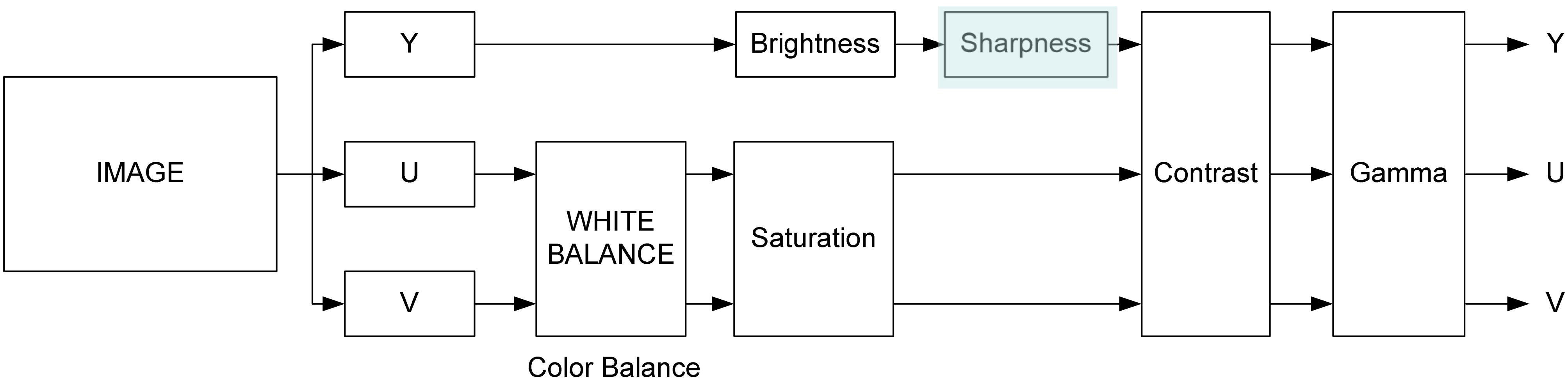
Sharpness

Sharpness (as a controllable parameter)

Examples.







Computer vision

- Broad definition of computer vision & artificial vision areas
- Some basic definitions
- The typical image processing pipeline
- Image parameters (intrinsic and adjustable)
- **Basic optics**
- Camera parameters and perspective views
- Examples of simple image processing tasks

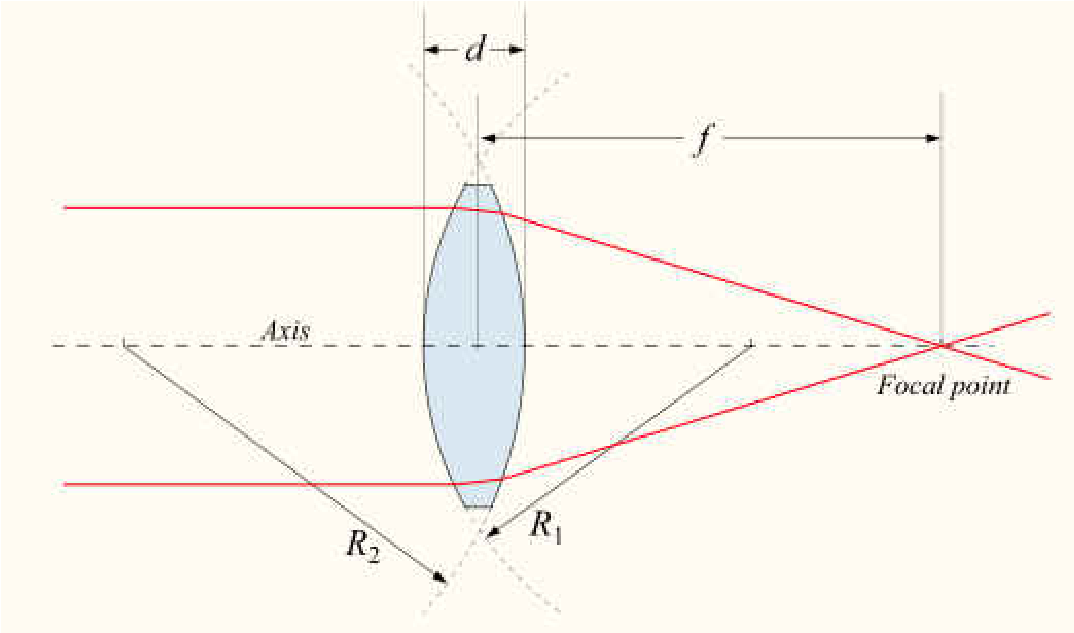
Optics basics

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, (e.g. 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.



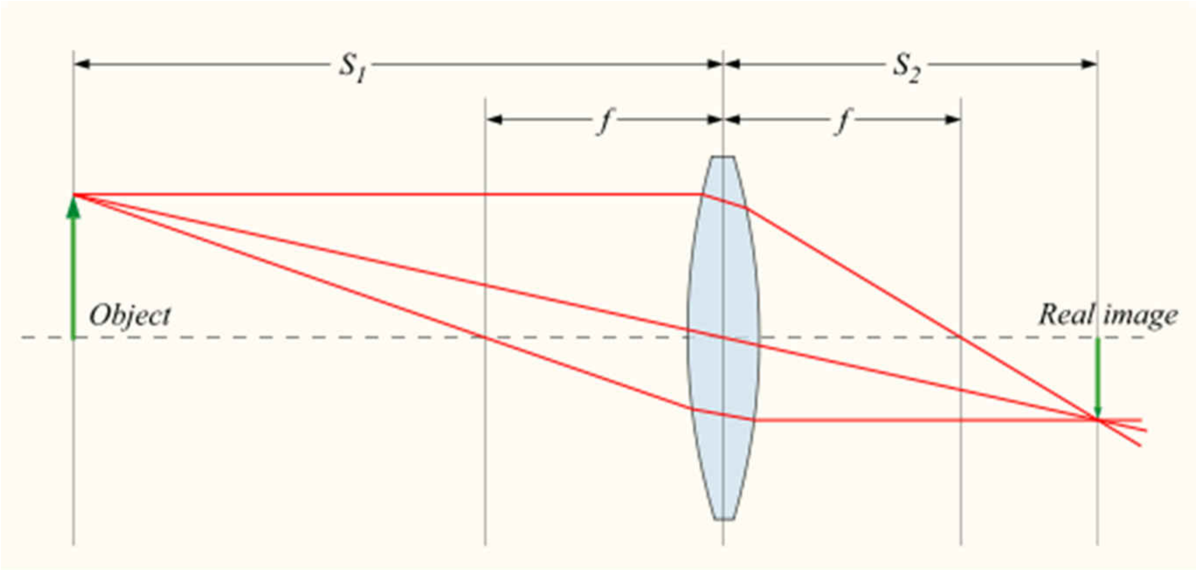
Positive (converging) lens

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



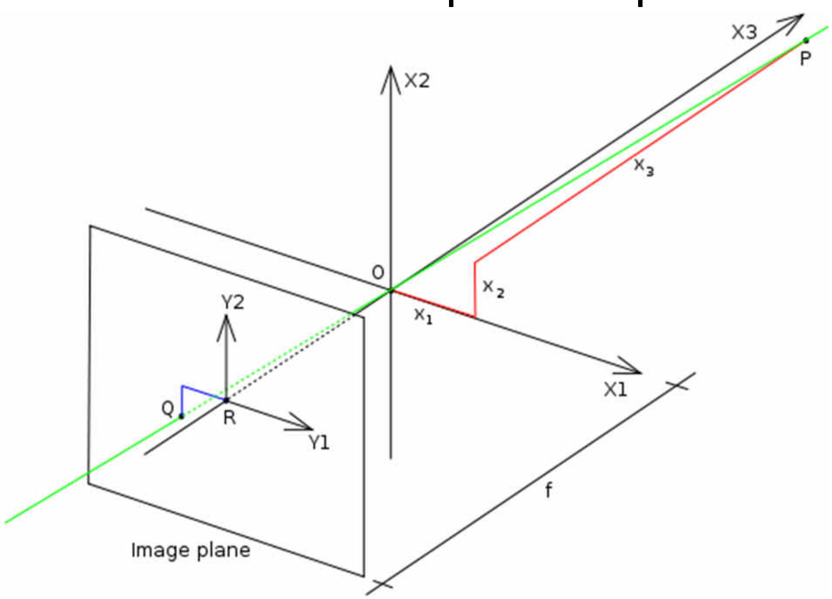
The pinhole model

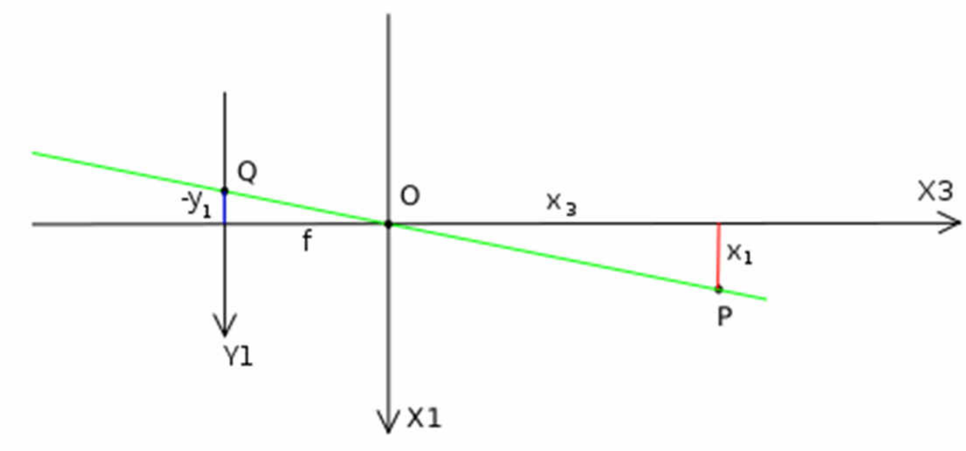
For most practical digital image processing tasks, the optical system can be approximated by the pinhole model.

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.





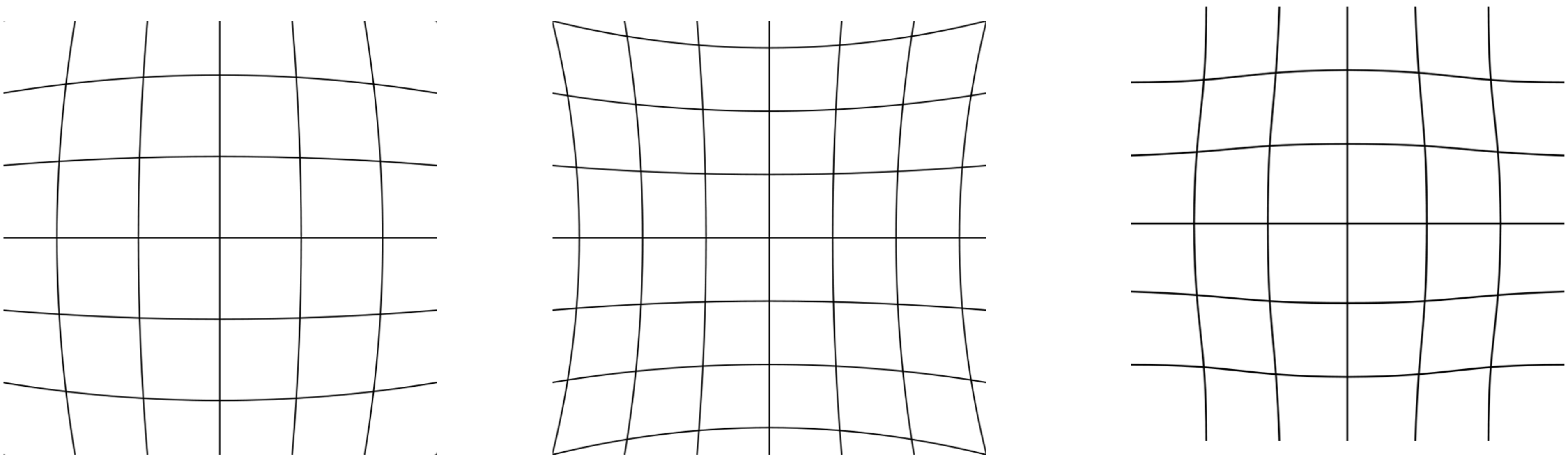
Lenses – Spherical Aberration

Spherical Aberration

Spherical Aberration is an image deformation resulting from the fact that most lenses have an external 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 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

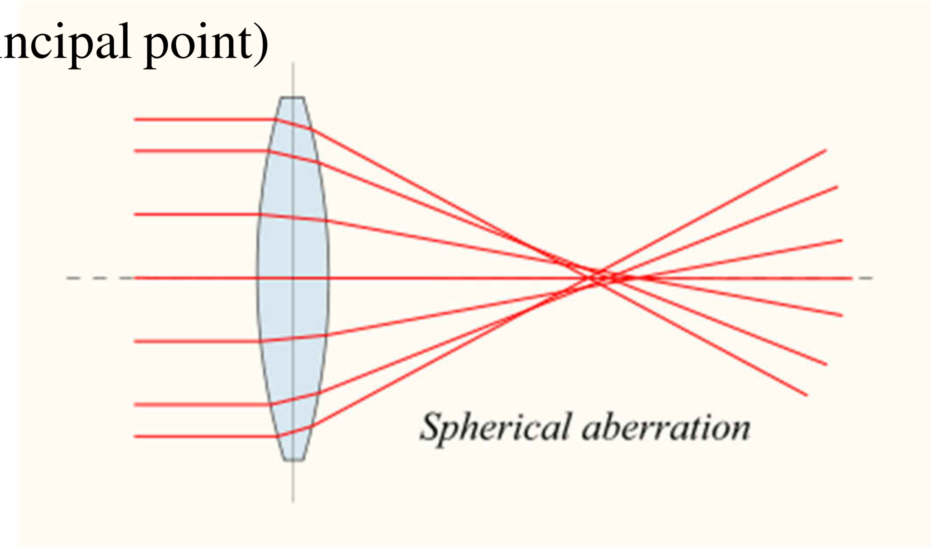
(x_d, y_d) - distorted image point as projected on image plane using specified lens

(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

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



Computer vision

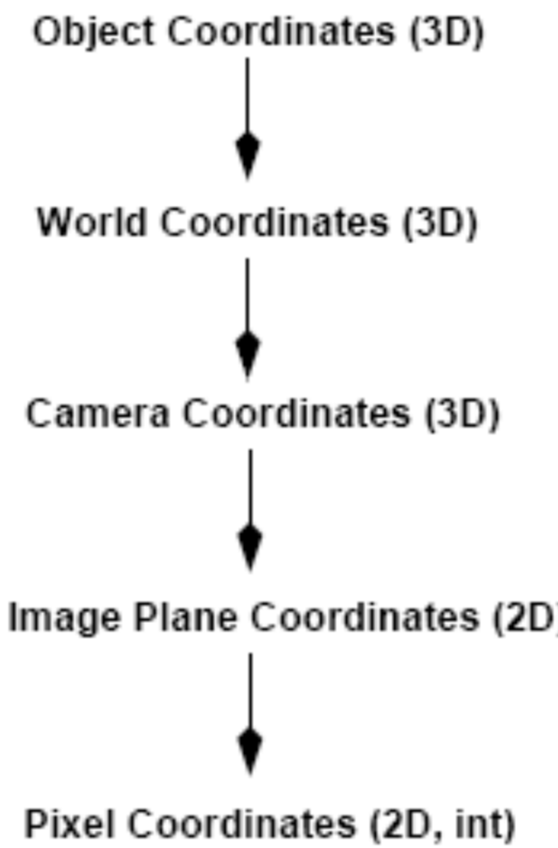
- Broad definition of computer vision & artificial vision areas
- Some basic definitions
- The typical image processing pipeline
- Image parameters (intrinsic and adjustable)
- Basic optics
- **Camera parameters and perspective views**
- Examples of simple image processing tasks

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.



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 (point $\mathbf{a}_n$ as the coordinates of the point to be projected, $\mathbf{c}_n$ as the pinhole coordinates and $\mathbf{d}_n$ as resulting rotated vector) can be obtained from:

$$\begin{bmatrix} \mathbf{d}_x \\ \mathbf{d}_y \\ \mathbf{d}_z \end{bmatrix} = \overbrace{\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\text{yaw}} \overbrace{\begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}}^{\text{pitch}} \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix}}^{\text{roll}} \left(\begin{bmatrix} \mathbf{a}_x \\ \mathbf{a}_y \\ \mathbf{a}_z \end{bmatrix} - \begin{bmatrix} \mathbf{c}_x \\ \mathbf{c}_y \\ \mathbf{c}_z \end{bmatrix} \right)$$

$$\begin{bmatrix} \mathbf{d}_x \\ \mathbf{d}_y \\ \mathbf{d}_z \end{bmatrix} = \begin{bmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{bmatrix} \left(\begin{bmatrix} \mathbf{a}_x \\ \mathbf{a}_y \\ \mathbf{a}_z \end{bmatrix} - \begin{bmatrix} \mathbf{c}_x \\ \mathbf{c}_y \\ \mathbf{c}_z \end{bmatrix} \right)$$

The translation vector ***Tv***, to apply to the point ***p*** can, on the other hand, can 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 we are allowed to combine a rotation matrix and a translation matrix into a single homogenous matrix.

Rotation Matrix

Translation Matrix

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

Camera parameters

Camera parameters are normally divided into two large groups:

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

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

Camera parameters

Camera <u>Intrinsic</u> parameters

Intrinsic parameters are defined as a matrix of the form

$$A = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

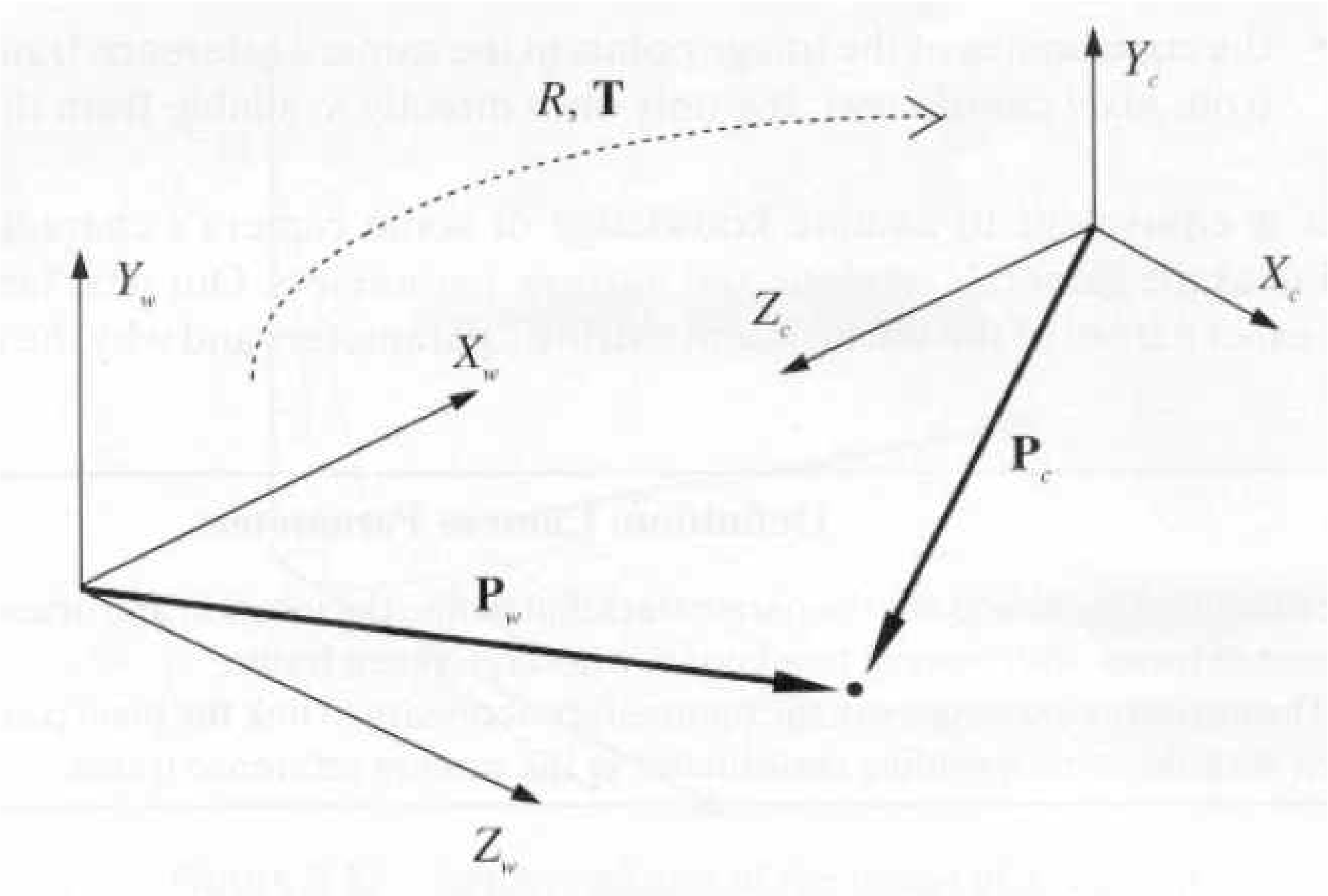
The intrinsic matrix containing 5 intrinsic parameters. These parameters include focal length, image format, and principal point. The parameters $\alpha_x = f \cdot m_x$ and $\alpha_y = f \cdot m_y$ represent focal length in terms of pixels, where m_x and m_y are the scale factors relating pixels to distance. γ represents the skew coefficient between the x and the y axis and is normally 0. u_0 and v_0 represent the principal point, which would be ideally in the center of the image.

Richard Hartley and Andrew Zisserman (2003). Multiple View Geometry in Computer Vision. Cambridge University Press. pp. 155–157. ISBN 0-521-54051-8

Camera parameters

Camera Extrinsic parameters

Extrinsic parameters denote the coordinate system transformations from 3D world coordinates to 3D camera coordinates.

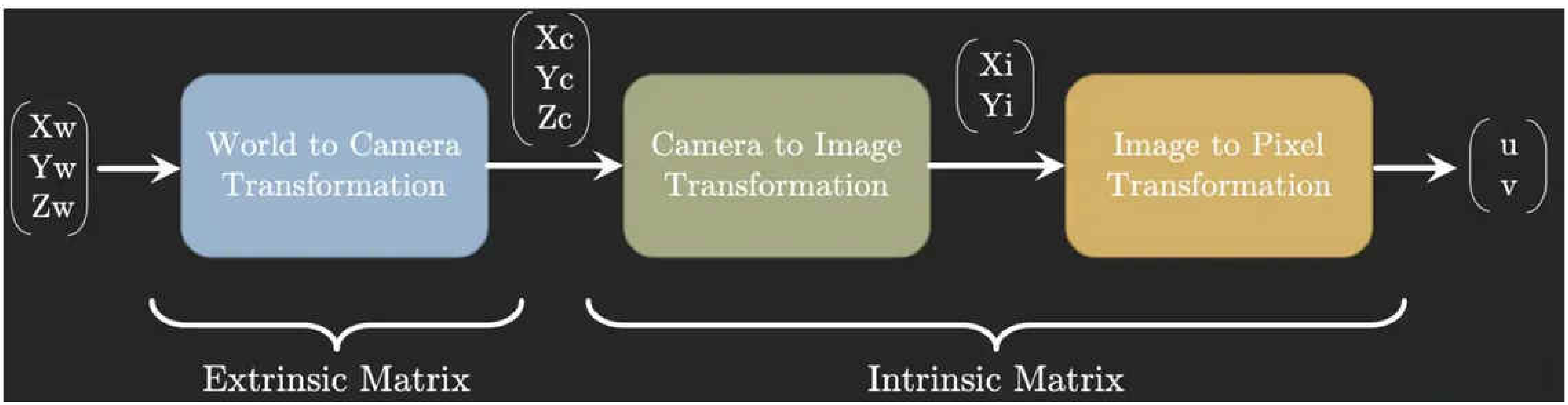


Camera parameters

Camera Extrinsic parameters

In other words, ***Extrinsic Parameters***, reflect the projection view of the camera taking into consideration its relative position to the world coordinates (translation operation) and its rotations when we consider the camera coordinate system three axis in relation to the world coordinate system three axis.

To note the **Extrinsic parameters** does not reflect the position of the camera in regard to the world coordinate system. It is, in fact, the position of the origin of the world coordinate system expressed in coordinates of the camera-centered coordinate system.



https://towardsdatascience.com

Parameters of the general model

Still referring to the pinhole camera model, a camera matrix can therefore be used to denote a general projective mapping from World coordinates.

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

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

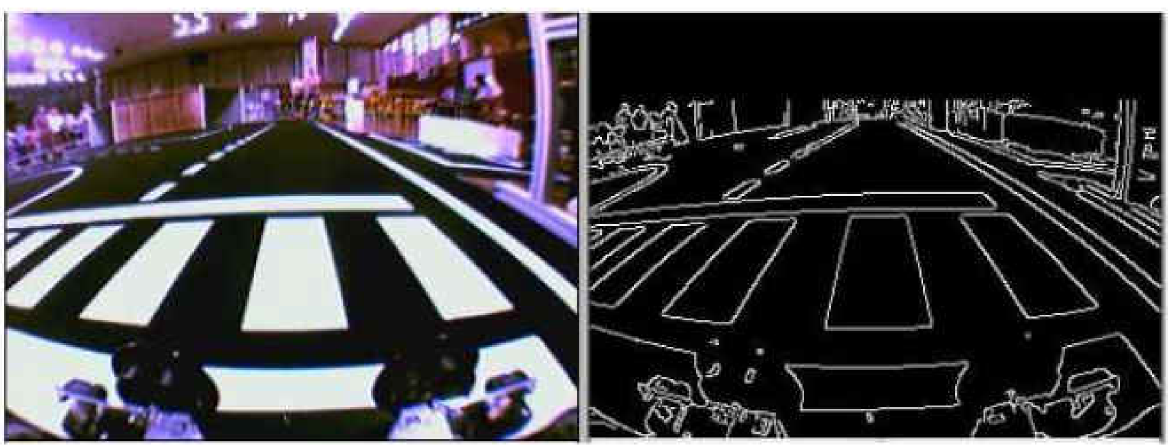
Computer vision

- Broad definition of computer vision & artificial vision areas
- Some basic definitions
- The typical image processing pipeline
- Image parameters (intrinsic and adjustable)
- Basic optics
- Camera parameters and perspective views
- **Examples of simple image processing tasks**

Examples of simple image processing tasks

Spatial filtering: applying a convolutional matrix

Low pass / High pass filtering

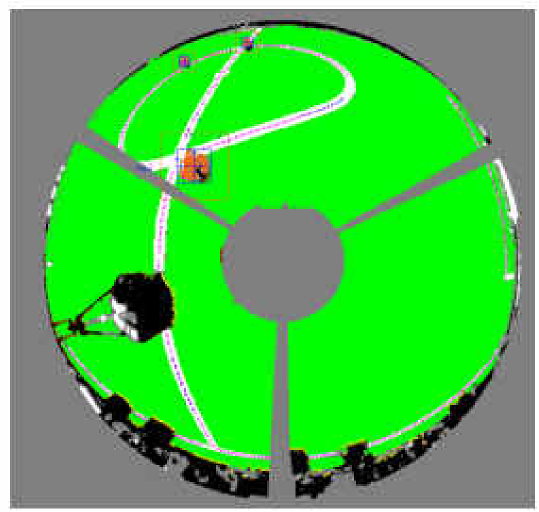


Edge detection (Canny, Laplace, Sobel, …

Examples of simple image processing tasks

Segmentation:

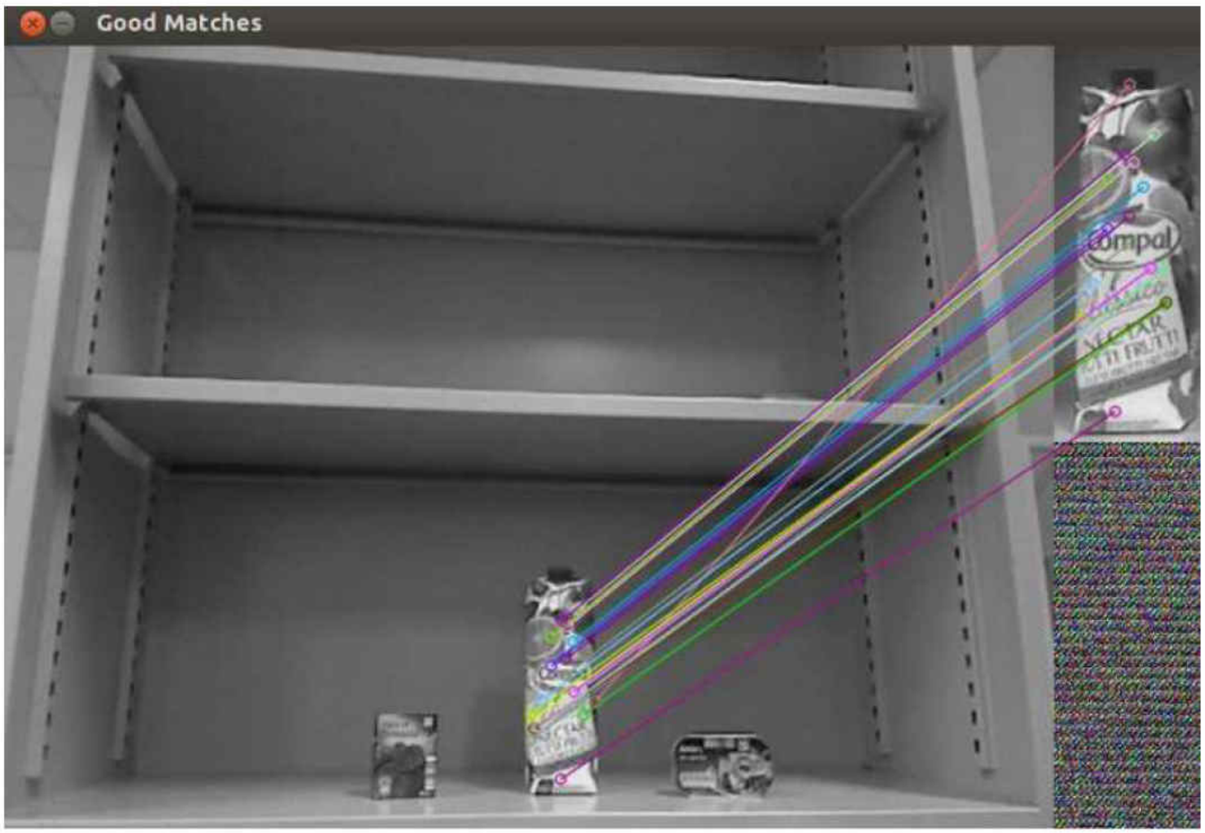




Other image processing tasks

Other more complex processing techniques allow more complex results but can not be covered here. Examples are:

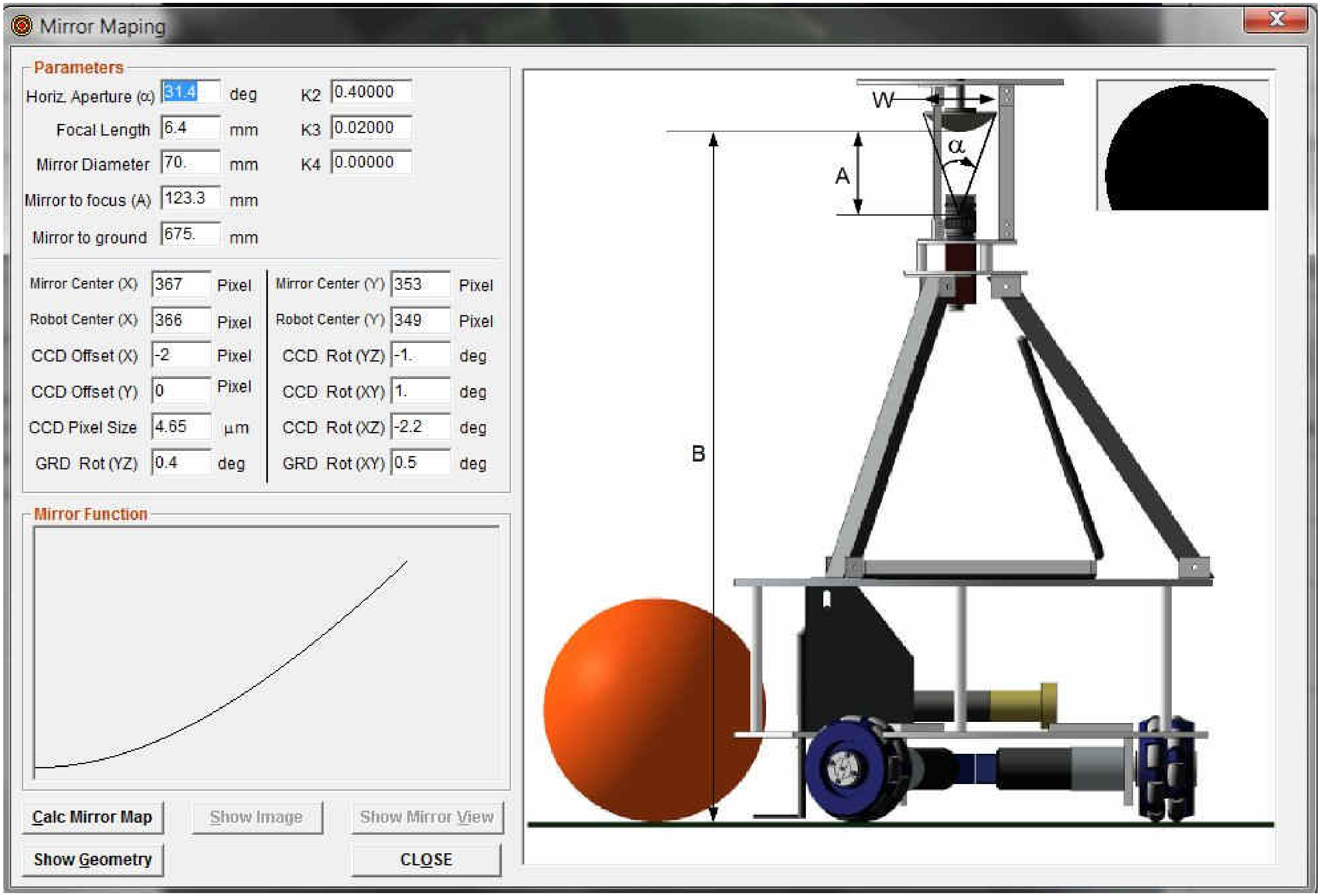
- Morphological operators
- Image descriptors
- Pattern recognition
- …



Example of use of image descriptors

Vision System – a special example

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



Vision System – a special example

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

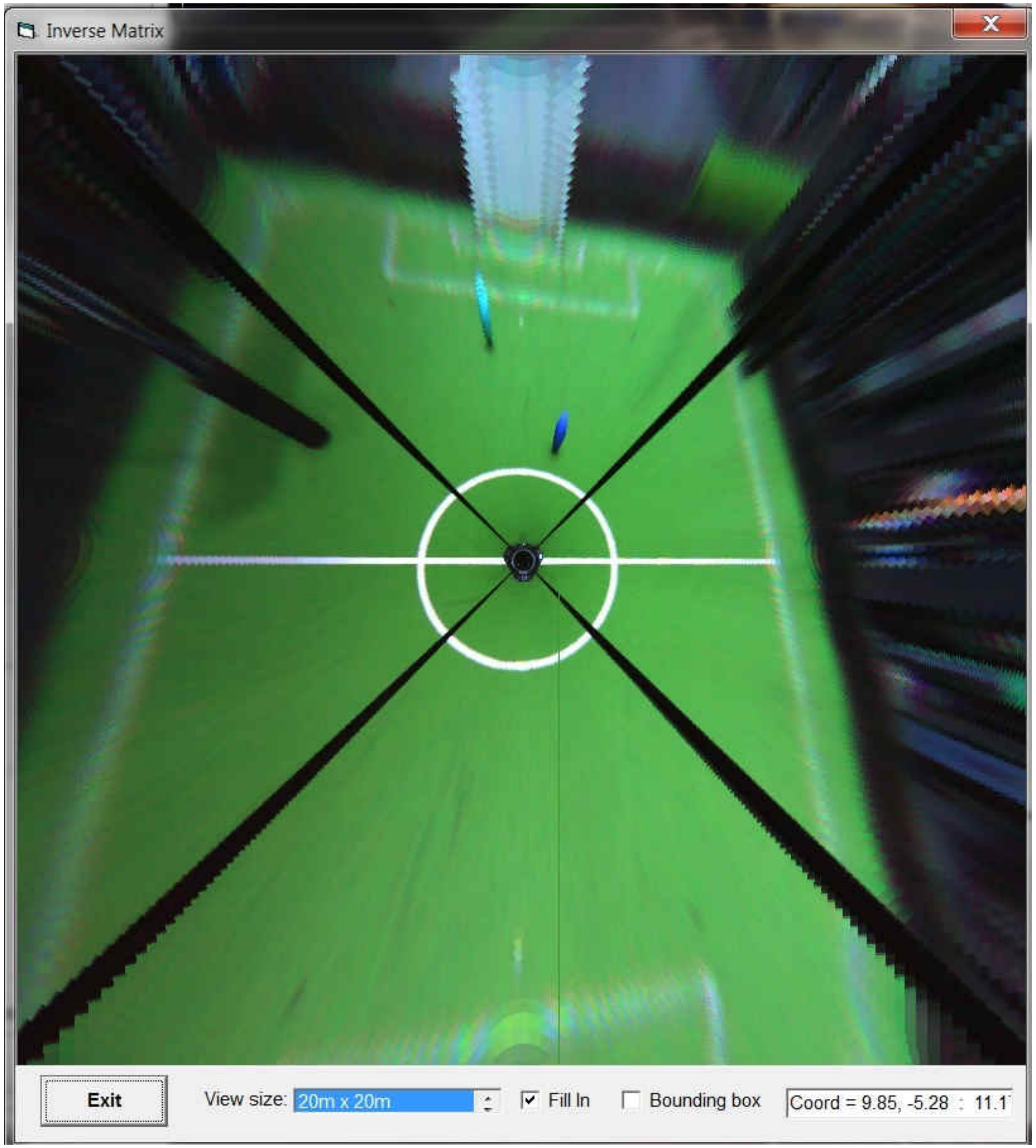
Vision System – a special example

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

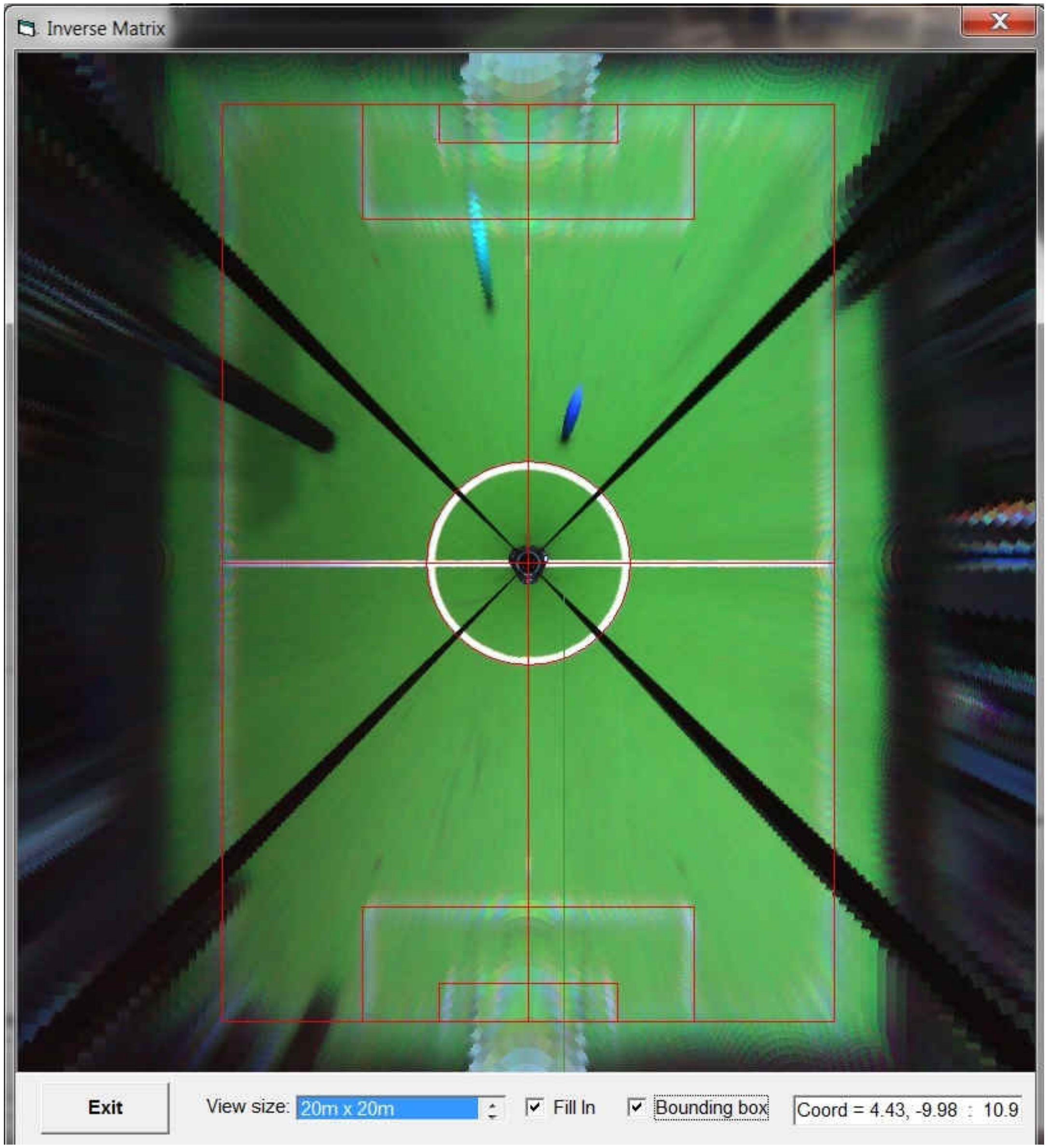
Vision System – a special example

- Setting all parameters to zero will normally result in a mapping distance matrix which is highly distorted
- the image on the right is an eagle eye view image obtained from the distance map in this case



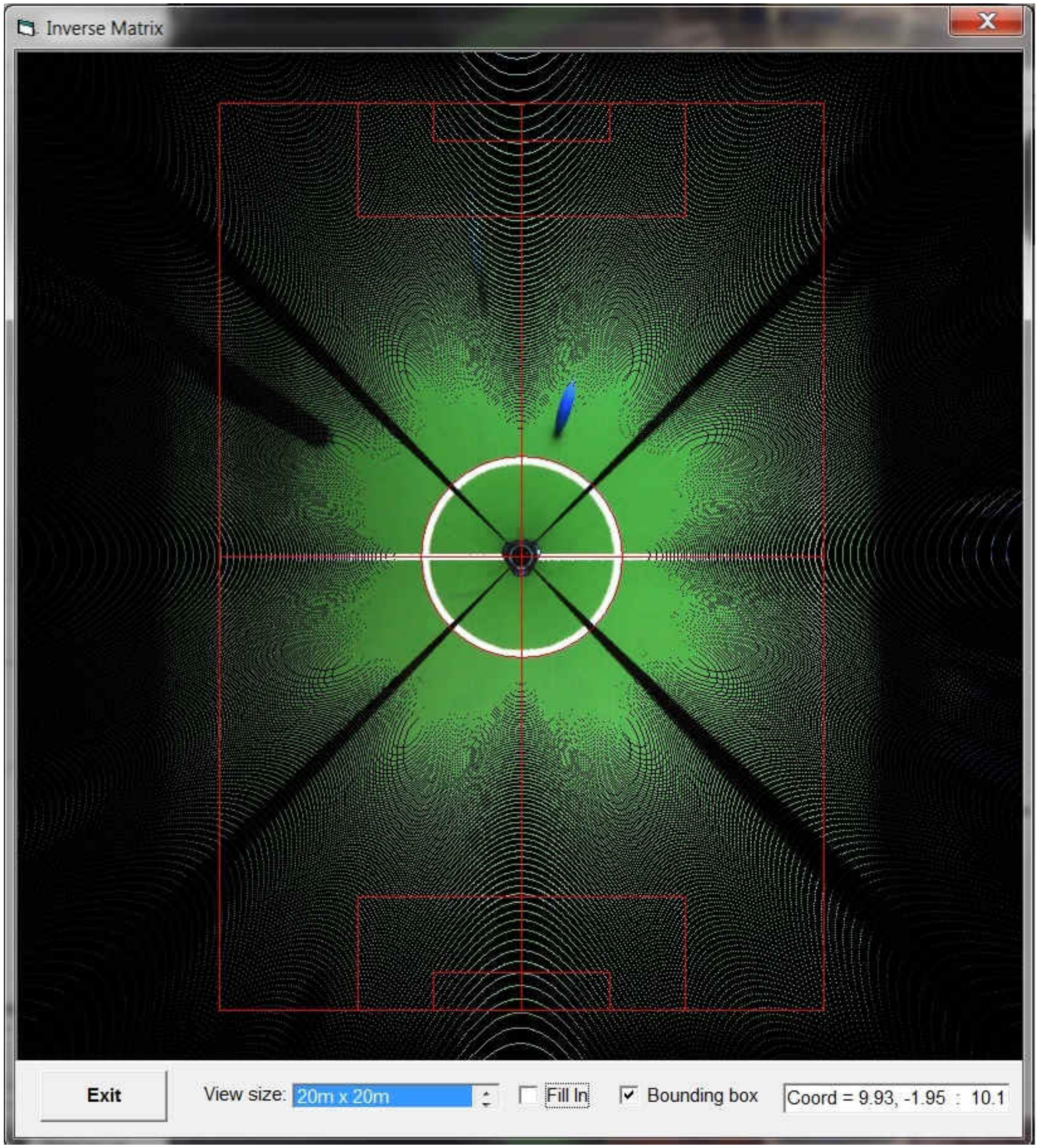
Vision System – a special example

- By adjusting the correction parameters one can obtain an “almost correct” distance matrix on the ground floor
- the image on the right is an eagle eye view image obtained from the distance map



Vision System – a special example

- The catadioptric effect highly affects the space resolution as we move from the center pixels to the peripheric ones
- the image on the previous slide is obtained by making a color integration among neighbor pixels according to its distance to the calculated resulting pixel coordinates



Some references

- 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
- Boris Peter Selby et al., "Patient positioning with X-ray detector self-calibration for image guided therapy", Australasian Physical & Engineering Science in Medicine, Vol.34, No.3, pages 391–400, 2011
- de Villiers, J. P.; Leuschner, F.W.; Geldenhuys, R. (17–19 November 2008). "Centi-pixel accurate real-time inverse distortion correction". 2008 International Symposium on Optomechatronic Technologies. SPIE. doi:10.1117/12.804771.
- Richard Hartley and Andrew Zisserman (2003). Multiple View Geometry in Computer Vision. Cambridge University Press. pp. 155–157. ISBN 0-521-54051-8.
- Sonka, M; Hlavac, V; Boyle, R (1995). Image Processing, Analysis & Machine Vision (2nd ed.). Chapman and Hall. pp. 14. ISBN 0-412-45570-6
- Ingrid Carlbom, Joseph Paciorek (1978). "Planar Geometric Projections and Viewing Transformations". ACM Computing Surveys 10 (4): 465–502. doi:10.1145/356744.356750
- Hecht, Eugene (1998). "5.7.6 The Camera". Optics (3rd ed.). ISBN 0-201-30425-2.