INTRO ROBOTICS AND VISION

Christoforos Kanellakis

LULEÅ UNIVERSITY OF TECHNOLOGY

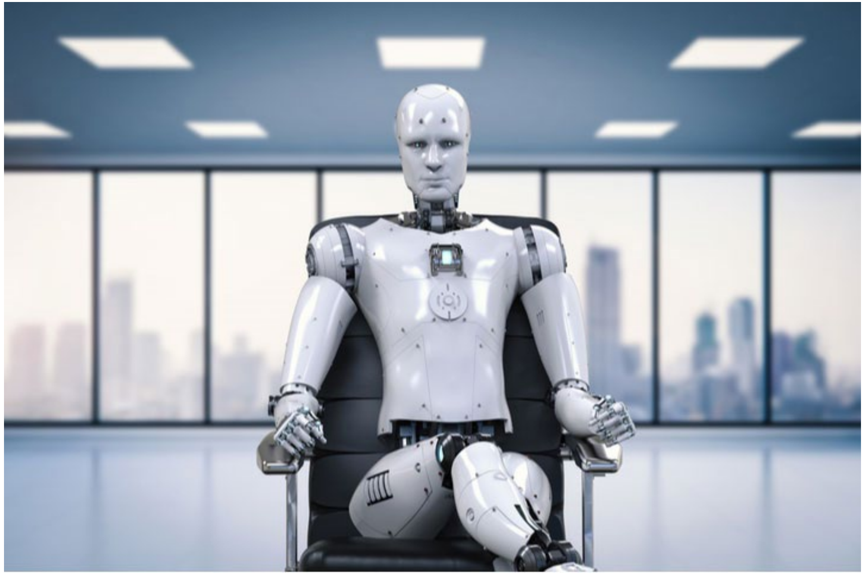
Learning Objectives

- Broaden horizons on the robots’ applications
- Knowledge of robots used in major industries
- Connection with visual sensors and properties
- Example on person detection

LULEÅ UNIVERSITY OF TECHNOLOGY

What is a Robot?

A goal-oriented machine that can sense, plan and act automatically



Why Robots?

- High precession
- Increased load capabilities
- Multi-tasking
- Repeat tasks with high accuracy
- Not affected by the monotony of tasks
- Robots can Collaborate with each other and to some extent they can collaborate with workers
- For humans to avoid hazardous areas
- Many robots can be reprogrammed to execute different tasks

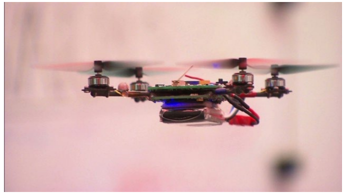
→

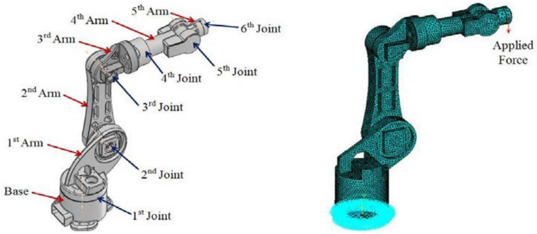
- Increased productivity
- Increased efficiency
- Cutting production costs
- Safer work environment

LULEÅ UNIVERSITY OF TECHNOLOGY

Robot types

- Robotic arms
- Mobile
- MAVs
- AUVs
- Legged
- Humanoid





(a) (b)

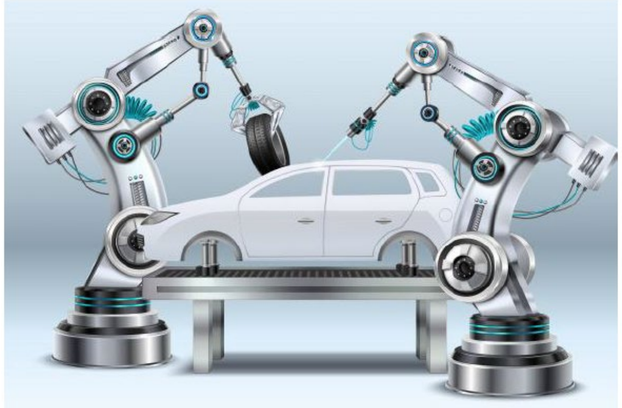
LULEÅ UNIVERSITY OF TECHNOLOGY

Where the Robots can be Used?

Industries that booming due to the increase in robot use:

- Manufacturing
- Agriculture
- Food Preparation
- Search and Rescue
- Pharmaceutical Industry
- Maintenance, Inspection and Repair
- Mining and Construction

Robot Sectors - Manufacturing



LULEÅ UNIVERSITY OF TECHNOLOGY

Robot Sectors - Agriculture

LULEÅ UNIVERSITY OF TECHNOLOGY

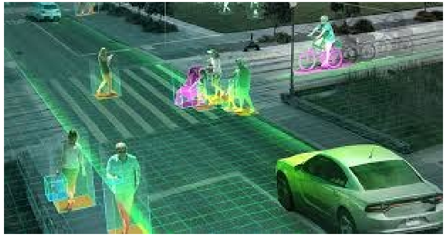
Robot Sectors – Food Industry

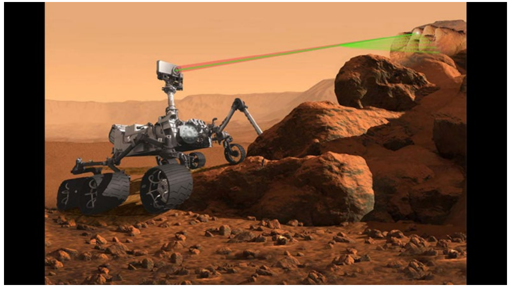
LULEÅ
UNIVERSITY
OF TECHNOLOGY

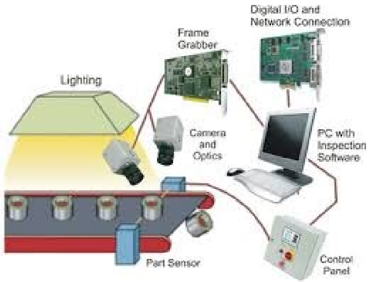
Robot Sectors – Search and Rescue

Delivery of time critical services

What is machine vision?







Automatic extraction of “meaningful” information from images and videos

LULEÅ UNIVERSITY OF TECHNOLOGY

Vision in Robotics

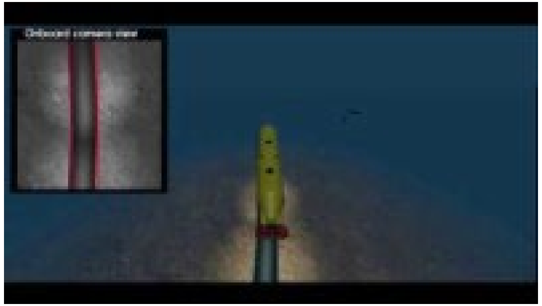
Focuses on computer vision methodologies and technical solutions applied to advanced robotic systems using conventional and intelligent paradigms

Vision in Robotics - Steps

- Object detection and tracking
- Feature extraction
- Object model acquisition

LULEÅ UNIVERSITY OF TECHNOLOGY

Focus on Vision



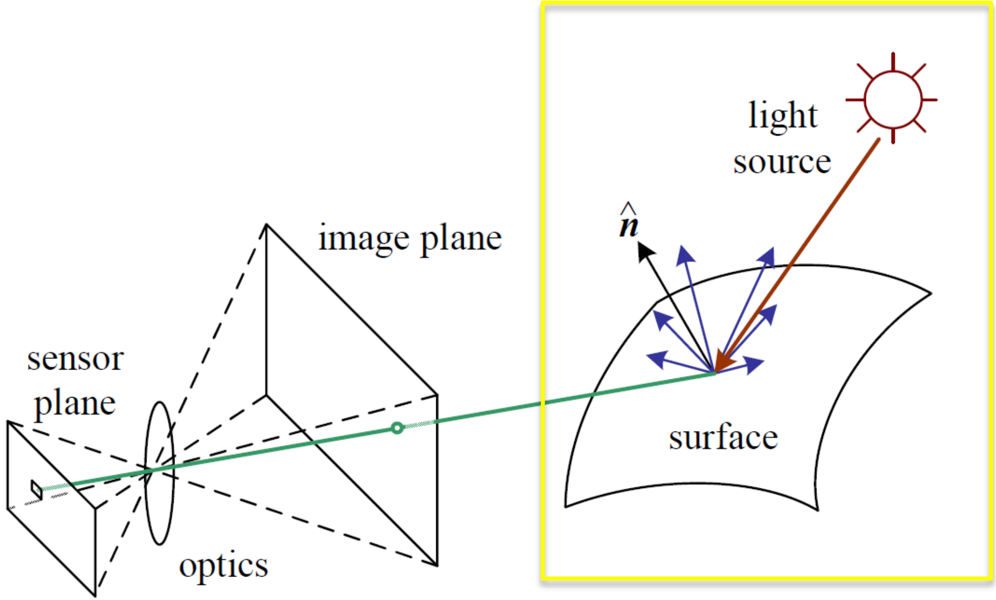
LULEÅ UNIVERSITY OF TECHNOLOGY

Vision in Robotics - Challenges

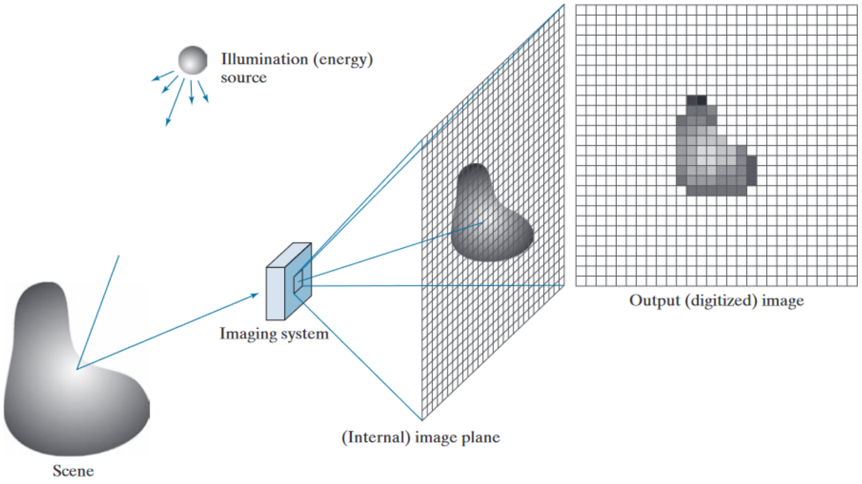
- Sensor Frame Rate
- Image/Depth processing
- Occlusion
- Viewpoint Handling
- Out-of-View events
- Clutter
- Camera Positioning
- Field of View

LULEÅ UNIVERSITY OF TECHNOLOGY

Image formation

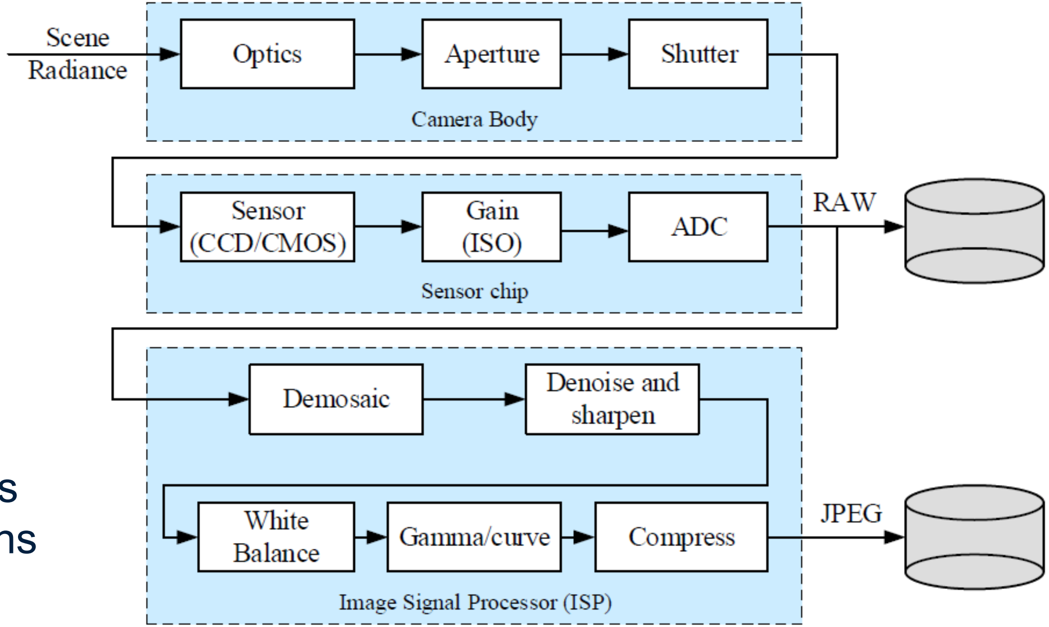


Photometric Image formation - The digital camera



LULEÅ UNIVERSITY OF TECHNOLOGY

The digital camera



- Sensor chip converts photons into electrons

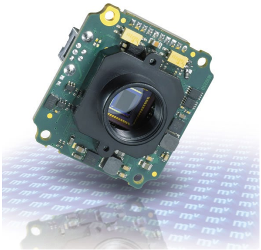
Why is hard? Interpreting different worlds

Jeong, J., Cho, Y., Shin, Y.-S., Roh, H., & Kim, A. (2019). Complex urban dataset with multi-level sensors from highly diverse urban environments. *The International Journal of Robotics Research*, *38*(6), 642–657. https://doi.org/10.1177/0278364919843996

239	245	255	126	7	24	0	0	0	0	0	8	9	0	0	0	0	3	0	0	0	2	0	192
254	221	34	0	16	0	0	11	8	0	0	8	5	0	0	2	0	3	0	0	0	2	0	192
208	17	0	21	0	3	9	3	0	1	1	0	0	0	1	4	0	3	0	0	0	2	0	192
99	0	0	10	0	8	0	0	9	0	0	0	9	9	0	0	0	3	0	0	0	2	0	192
192	195	196	188	190	194	192	201	191	93	8	0	4	0	0	8	0	3	0	0	0	2	0	192
255	255	255	255	255	255	255	255	255	126	10	0	7	4	0	1	0	3	0	0	0	2	0	192
128	128	128	128	128	128	128	128	123	62	6	0	1	0	0	3	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	8	5	2	0	0	0	0	0	0	3	0	0	0	2	0	192
1	1	1	1	1	1	1	1	0	0	0	0	2	5	1	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	9	2	0	1	3	1	0	0	0	3	0	0	0	2	0	192
2	2	2	2	2	2	2	2	2	0	0	3	0	0	0	8	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	6	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	1	4	3	0	0	4	2	0	0	3	0	0	0	2	0	192
7	0	0	5	0	2	17	0	1	0	9	1	0	0	5	0	0	3	0	0	0	2	0	192
0	4	0	6	5	0	0	0	0	16	0	0	0	1	8	3	0	3	0	0	0	2	0	192
3	15	0	0	9	11	15	0	57	123	18	3	2	1	0	0	0	3	0	0	0	2	0	192
0	0	0	11	0	0	0	2	68	182	10	0	3	0	0	3	0	3	0	0	0	2	0	192
201	48	1	117	173	155	102	29	57	202	0	0	2	1	0	5	0	3	0	0	0	2	0	192
240	3	107	212	92	124	246	59	66	216	0	1	5	0	0	0	0	3	0	0	0	2	0	192
41	14	161	86	0	4	189	85	52	192	0	0	0	0	2	0	0	3	0	0	0	2	0	192
3	4	188	58	5	1	202	54	68	195	0	6	3	0	5	0	0	3	0	0	0	2	0	192
0	0	170	122	2	19	243	65	64	201	0	0	3	1	0	10	0	3	0	0	0	2	0	192
192	99	38	243	205	197	242	59	70	175	11	9	9	0	0	0	0	3	0	0	0	2	0	192
67	42	34	46	67	36	222	42	8	69	0	0	0	0	2	0	0	3	0	0	0	2	0	192
0	0	106	221	122	166	204	4	0	10	0	0	0	0	6	0	0	3	0	0	0	2	0	192
10	0	0	74	135	119	20	0	9	0	9	3	0	0	5	0	0	3	0	0	0	2	0	192
0	18	0	0	0	0	0	14	0	0	5	0	0	1	5	1	0	3	0	0	0	2	0	192
0	6	10	0	21	0	0	0	14	0	16	0	0	2	0	0	0	3	0	0	0	2	0	192
11	0	11	0	0	6	1	2	0	0	4	0	0	1	0	2	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	192
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12	0	7	7	0	16	188
190	190	190	190	190	190	190	190	190	190	190	190	190	190	190	190	184	198	177	182	201	196	179	255

LULEÅ UNIVERSITY OF TECHNOLOGY

Visual Servoing – Typical Sensos

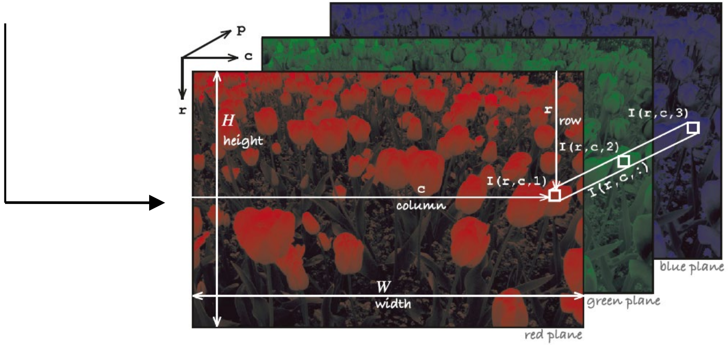


LULEÅ UNIVERSITY OF TECHNOLOGY

Color spaces

RGB (3 channels)

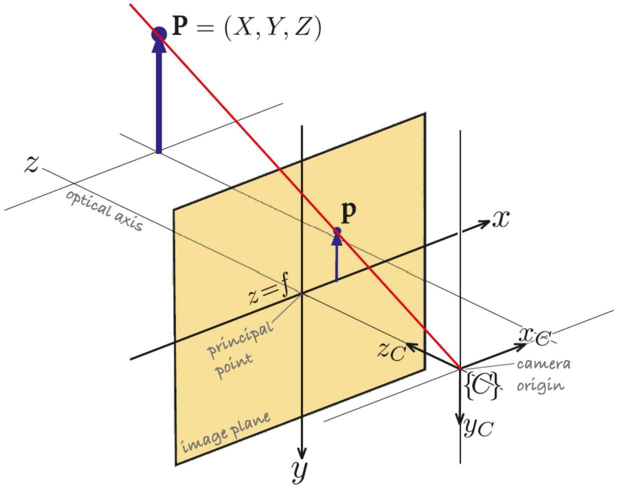
greyscale (1 channel)



other spaces
CMY
YIQ
HSI
HSV
HLS
LAB
HMMD

LULEÅ UNIVERSITY OF TECHNOLOGY

Perspective camera modelling



World coordinates: $\mathbf{P} = (X, Y, Z)$

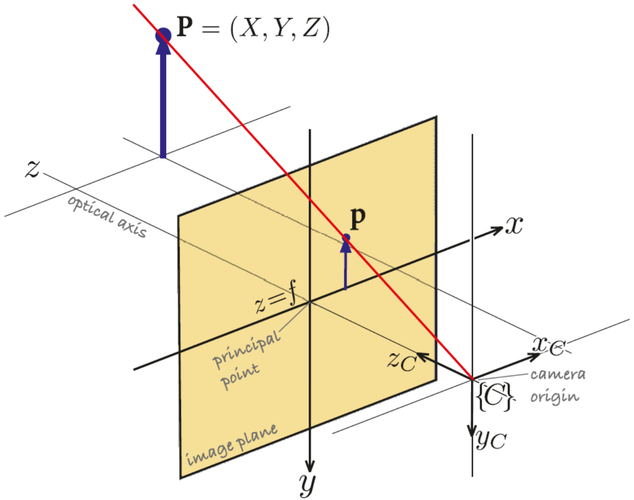
Image coordinates: $\mathbf{p} = (x, y)$

$\mathbf{P}$ is projected to $\mathbf{p}$ through perspective projection

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

LULEÅ UNIVERSITY OF TECHNOLOGY

Perspective camera modelling



Homogeneous world coordinates: ${}^{C}\tilde{\boldsymbol{P}} = (X, Y, Z, 1)^T$

Homogeneous image coordinates: $\tilde{\boldsymbol{p}} = (\tilde{x}, \tilde{y}, \tilde{z})$

$$\downarrow$$

$$\tilde{x} = fX,\ \tilde{y} = fY,\ \tilde{z} = Z$$

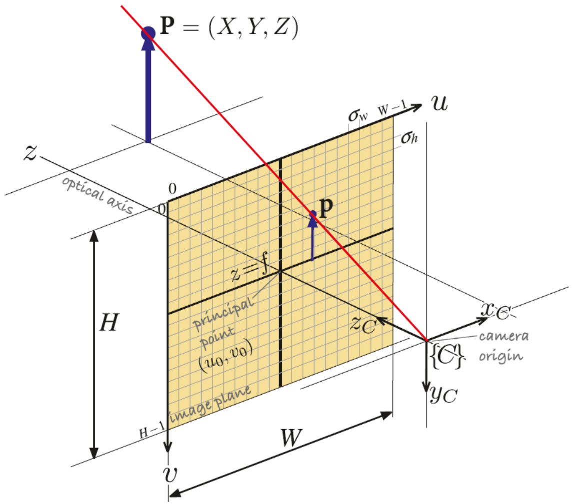
$$\downarrow$$

$$\tilde{\boldsymbol{p}} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} {}^{C}\tilde{\boldsymbol{P}}$$

camera matrix C

Nonhomogeneous image coordinates: $x = \dfrac{\tilde{x}}{\tilde{z}},\ y = \dfrac{\tilde{y}}{\tilde{z}}$

Discrete image plane



pixel coordinates: $u = \dfrac{x}{\rho_w} + u_0,\ v = \dfrac{y}{\rho_h} + v_0$

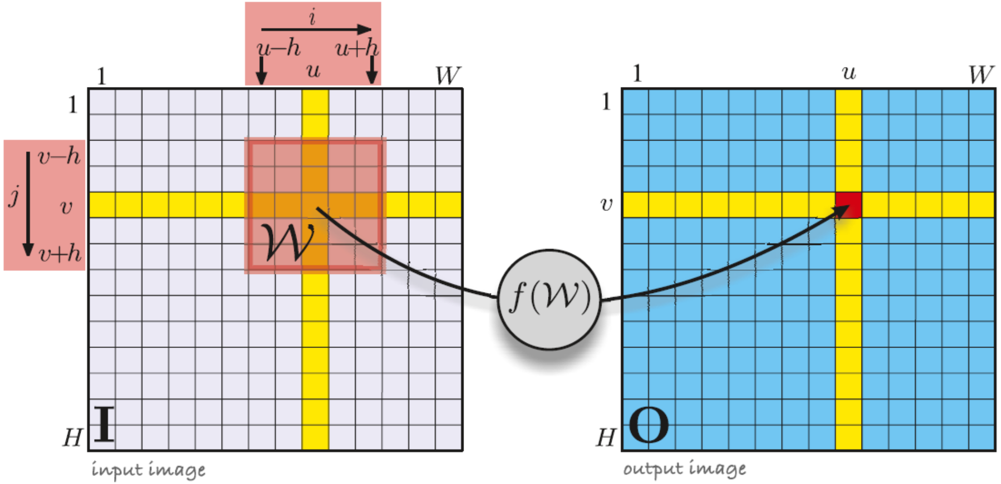
Homogeneous world coordinates: ${}^C\tilde{\boldsymbol{P}} = (X, Y, Z, 1)^T$

Homogeneous pixel coordinates: $\tilde{\boldsymbol{p}} = (\tilde{u}, \tilde{v}, \tilde{w})$

$$\tilde{\boldsymbol{p}} = \underbrace{\begin{pmatrix} 1/\rho_w & 0 & u_0 \\ 0 & 1/\rho_h & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{K} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} {}^C\tilde{\boldsymbol{P}}$$

Nonhomogeneous pixel coordinates: $u = \dfrac{\tilde{u}}{\tilde{w}},\ v = \dfrac{\tilde{v}}{\tilde{w}}$

Spatial Operations on Images



correlation $\longrightarrow$ $\mathbf{O} = \boldsymbol{K} \otimes \mathbf{I}$ $\longrightarrow$ $\mathbf{O}[u,v] = \sum_{(i,j)\in\mathcal{W}} \boldsymbol{I}[u+i, v+j]\boldsymbol{K}[i,j], \ \forall (u,v) \in \mathbf{I}$

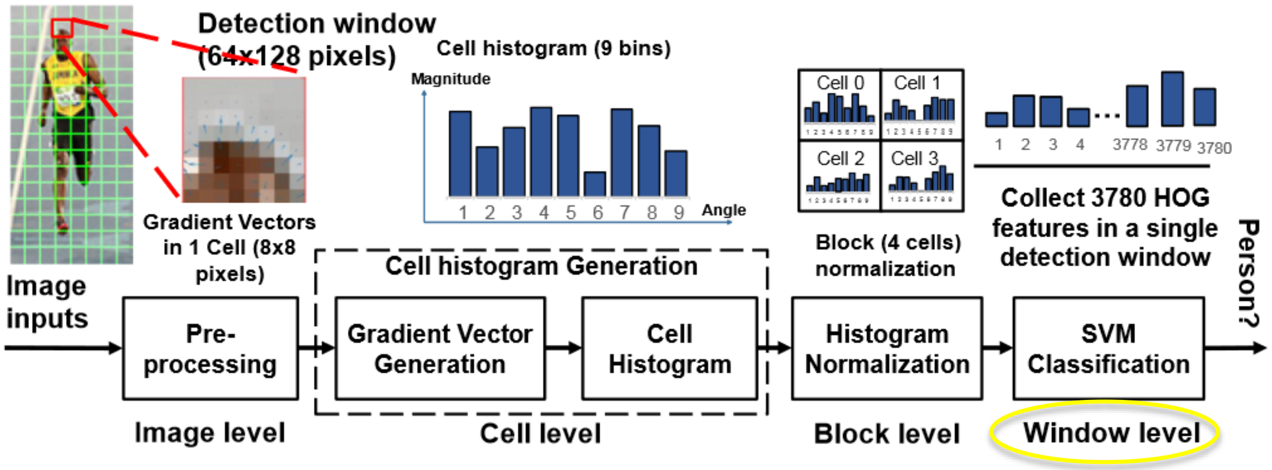
convolution $\longrightarrow$ $\mathbf{O} = \boldsymbol{K} * \mathbf{I}$ $\longrightarrow$ $\mathbf{O}[u,v] = \sum_{(i,j)\in\mathcal{W}} \boldsymbol{I}[u-i, v-j]\boldsymbol{K}[i,j], \ \forall (u,v) \in \mathbf{I}$

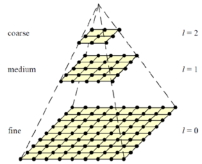
LULEÅ UNIVERSITY OF TECHNOLOGY

Example on Human Detection

Sliding window method

Image Pyramids





[1] Suleiman, A., & Sze, V. (2014, October). Energy-efficient HOG-based object detection at 1080HD 60 fps with multi-scale support. In 2014 IEEE Workshop on Signal Processing Systems (SiPS) (pp. 1-6). IEEE.

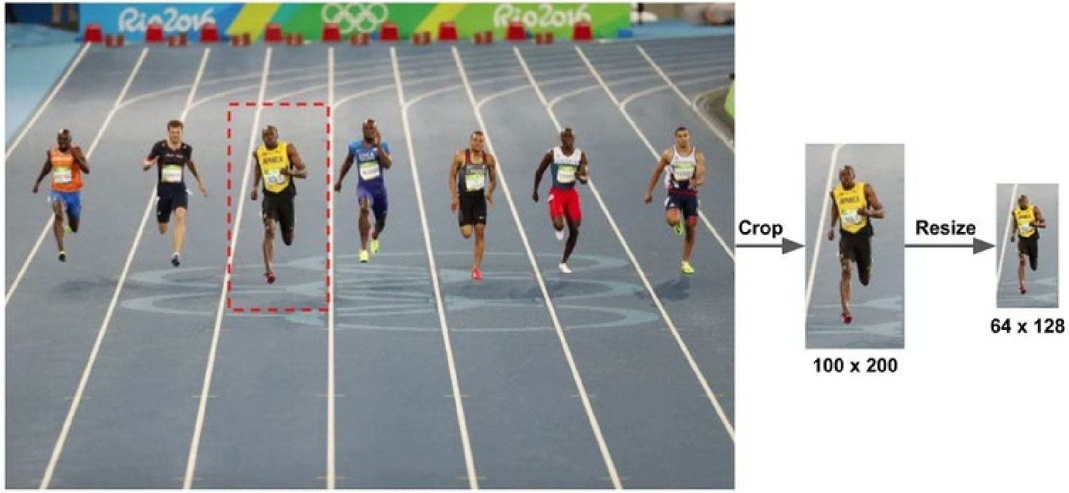
[2] Nguyen, N. D., Bui, D. H., & Tran, X. T. (2019, November). A novel hardware architecture for human detection using HOG-SVM co-optimization. In 2019 IEEE Asia Pacific conference on circuits and systems (APCCAS) (pp. 33-36). IEEE.

LULEÅ UNIVERSITY OF TECHNOLOGY

Histogram of Oriented Gradients (HOG) Descriptor

Creates a vector with histograms of directions of gradients

Step 1: Create patch of 64x128



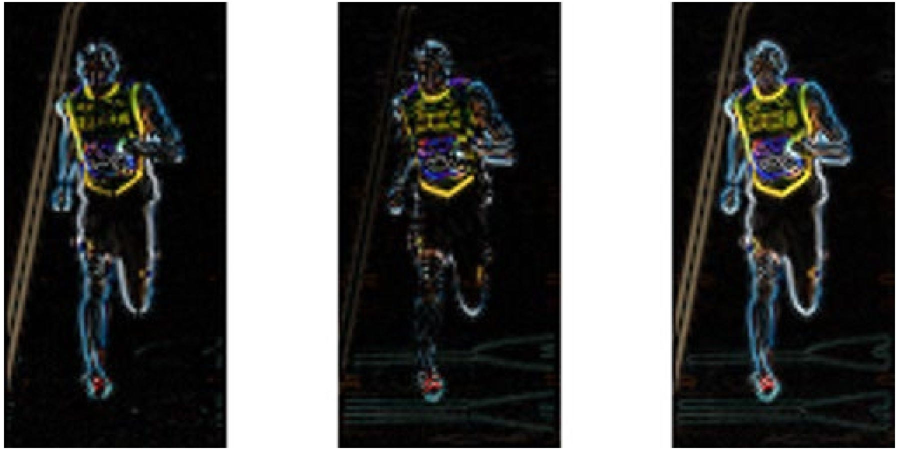
Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.
https://learnopencv.com/histogram-of-oriented-gradients/

Histogram of Oriented Gradients (HOG) Descriptor

Creates a vector with histograms of directions of gradients

Step 1: Create patch of 64x128

Step 2: Calculate image gradients



$$\nabla \boldsymbol{f} = \left[\frac{\partial \boldsymbol{f}}{\partial x}, \frac{\partial \boldsymbol{f}}{\partial y}\right]$$

$$M(x,y) = \|\nabla f(x,y)\| = \sqrt{g_x^2(x,y) + g_y^2(x,y)}$$

gradient magnitude

$$\alpha(x,y) = \tan^{-1}\left[\frac{g_y(x,y)}{g_x(x,y)}\right]$$

gradient direction

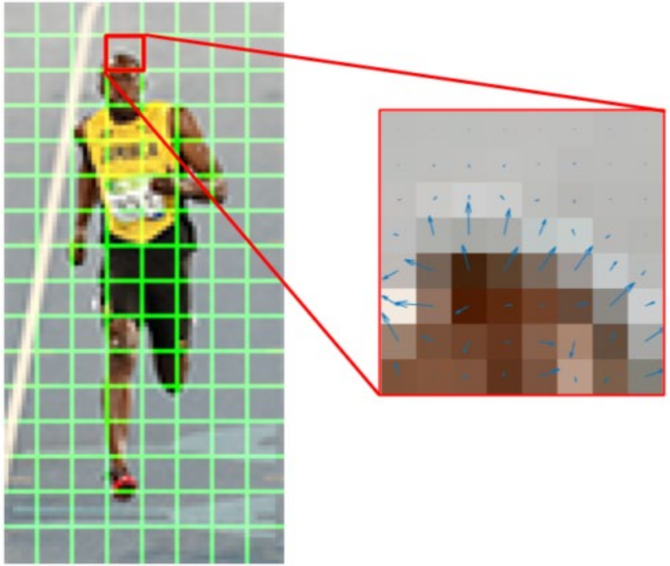
Histogram of Oriented Gradients (HOG) Descriptor

Creates a vector with histograms of directions of gradients

Step 1: Create patch of 64x128

Step 2: Calculate image gradients

Step 3: Calculate histogram of unsigned gradients in 8x8 cells



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Gradient Magnitude

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

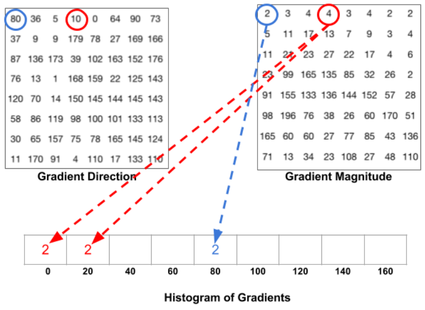
Histogram of Oriented Gradients (HOG) Descriptor

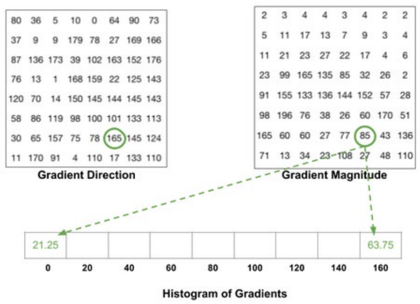
Creates a vector with histograms of directions of gradients

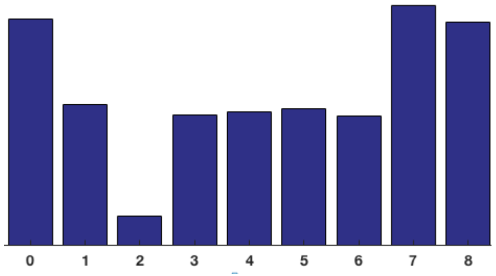
Step 1: Create patch of 64x128

Step 2: Calculate image gradients

Step 3: Calculate histogram of gradients in 8x8 cells







Histogram of Oriented Gradients (HOG) Descriptor

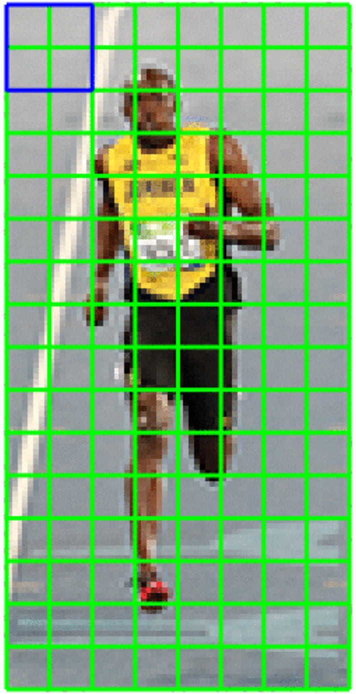
Creates a vector with histograms of directions of gradients

Step 1: Create patch of 64x128

Step 2: Calculate image gradients

Step 3: Calculate histogram of gradients in 8x8 cells

Step 4: 16x16 Block Normalization (robustness to illumination variations)



LULEÅ UNIVERSITY OF TECHNOLOGY

Histogram of Oriented Gradients (HOG) Descriptor

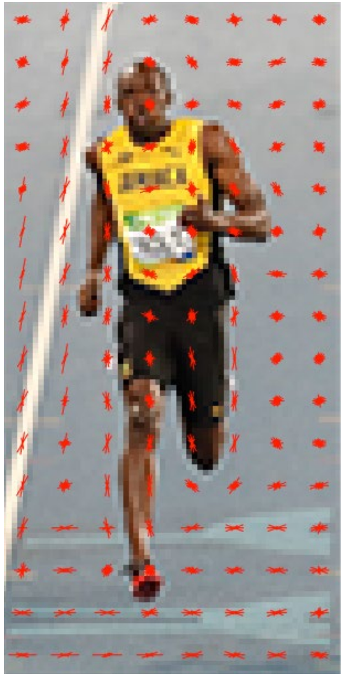
Creates a vector with histograms of directions of gradients

Step 1: Create patch of 64x128

Step 2: Calculate image gradients

Step 3: Calculate histogram of gradients in 8x8 cells

Step 4: 16x16 Block Normalization (robustness to illumination variations)

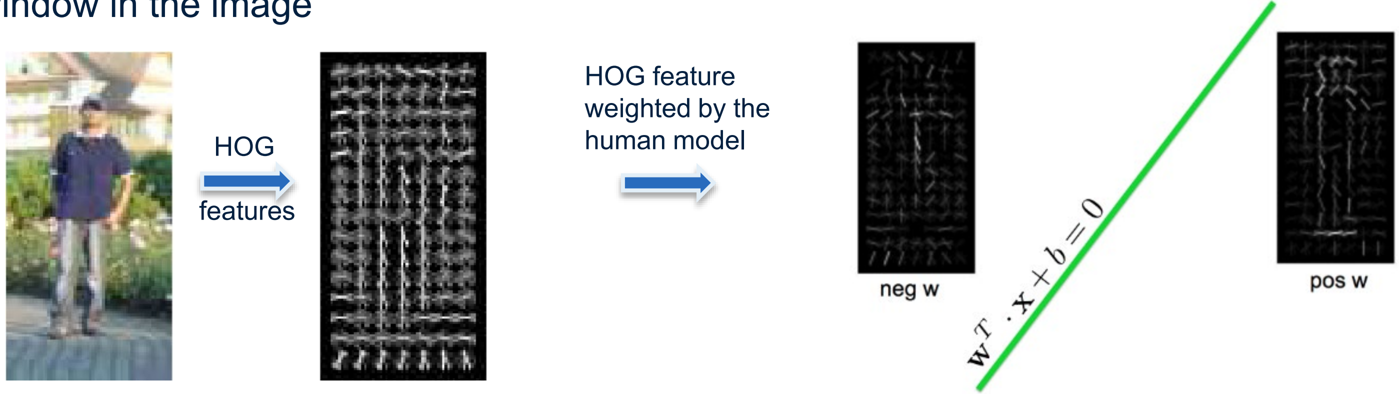


Step 5: concatenate 36x1 vectors of 7x15=105 block positions (36x105=3780 vector size)

LULEÅ UNIVERSITY OF TECHNOLOGY

SVM Classification

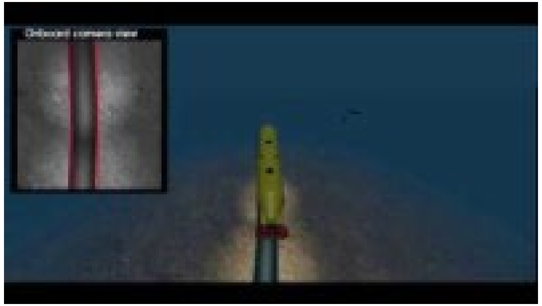
- Train a window classifier
- Use the trained classifier to predict the object class in each window in the image



Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.

Ahmed, A. H., Kpalma, K., & Guedi, A. O. (2017, December). Human detection using HOG-SVM, mixture of Gaussian and background contours subtraction. In 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 334-338). IEEE.

Focus on Vision



LULEÅ
UNIVERSITY
OF TECHNOLOGY

Matlab Time

Break and meet during the lab time slot

```matlab
%% Matlab code D7062E

I=imread('cameraman.tif');
%%
figure,imshow(I)
%%
I_d=im2double(I);
%%
max(I(:))
%%
max(I_d(:))
%%
size(I)
%%
C=imread('autumn.tif'); %24-bit
%%
figure,imshow(C)
%%
size(C)
%%
C(100,200,2)
%%
C(100,200,:)
%%
impixel(C,200,100)
%%
imfinfo('cameraman.tif')
%% Data conversion
A =[-8.0000 4.0000 0 0.5000]
B = uint8(A)
%%
Cgr=rgb2gray(C);
%%
figure,imshow(Cgr)
%%
Cdouble=double(Cgr);
%%
```

LULEÅ UNIVERSITY OF TECHNOLOGY

LULEÅ
UNIVERSITY
OF TECHNOLOGY