



Advanced Vision Systems

AGH UST International Course

Tomasz Kryjak, PhD, Mateusz Wąsala, MSc



Copyright © 2023

Tomasz Kryjak

Mateusz Wąsala

EMBEDDED VISION SYSTEMS GROUP

VISION SYSTEMS LABORATORY

WEAiIB, AGH UNIVERSITY OF KRAKOW

Third edition, revised and corrected, Krakow, February 2024

Contents

I	Laboratory 1	
1	Vision algorithms in Python 3.X – introduction	7
1.1	Programming software	7
1.2	Python modules used in image processing	7
1.3	Input/output operations	8
1.3.1	Reading, displaying and saving an image using OpenCV	8
1.3.2	Loading, displaying and saving an image using the <i>Matplotlib</i> module	9
1.4	Colour space conversion	10
1.4.1	OpenCV	10
1.4.2	Matplotlib	10
1.5	Scaling, rescaling with OpenCV	11
1.6	Arithmetic operations: addition, subtraction, multiplication, difference module	11
1.7	Histogram calculation	12
1.8	Histogram equalisation	12
1.9	Filtration	13
II	Laboratory 2	
2	Detection of foreground moving objects	17
2.1	Image sequence loading	17
2.2	Frame subtraction and Binarization	18

2.3	Morphological operations	19
2.4	Labeling and simple analysis	19
2.5	Evaluation of foreground object detection results	20
2.6	Example results of the algorithm for foreground moving object detection	21

III

Laboratory 3

3	Foreground object segmentation	25
3.1	Main objectives	25
3.2	Theory of foreground object segmentation	25
3.3	Methods based on frame buffer	26
3.4	Approximation of mean and median (so-called sigma-delta)	27
3.5	Updating policy	28
3.6	OpenCV – GMM/MOG	28
3.7	OpenCV – KNN	28
3.8	Example results of the algorithm for foreground object segmentation	29
3.9	Using a neural network to segment foreground objects.	31
3.10	Additional exercises	31
3.10.1	Initialization of the background model in the presence of foreground objects on the scene	31
3.10.2	Implementation of VIBE and PBAS methods	32

Laboratory 1

1	Vision algorithms in Python 3.X – introduction	7
1.1	Programming software	
1.2	Python modules used in image processing	
1.3	Input/output operations	
1.4	Colour space conversion	
1.5	Scaling, rescaling with OpenCV	
1.6	Arithmetic operations: addition, subtraction, multiplication, difference module	
1.7	Histogram calculation	
1.8	Histogram equalisation	
1.9	Filtration	

1. Vision algorithms in Python 3.X – introduction

The exercise will present/mention basic information related to the use of the Python language to perform image and video processing operations and image analysis.

1.1 Programming software

Before starting the exercises, **please create** your own working directory in the place given by the tutor. The name of the directory should be associated with your name – the recommended form is LastName_FirstName without no special characters.

The Python programming environment – Visual Studio Code (VSCoDe) – can be used to complete the exercises. Visual Studio Code is a free, lightweight, intuitive and easy to use code editor. It is a universal software for any programming language. Configuration of the editor for exercises comes down only to choosing the appropriate interpreter, or more precisely the appropriate virtual environment in the Python language.

1.2 Python modules used in image processing

Python does not have a native array type that allows to perform elementwise operations between two containers (in a convenient and efficient way). For operations on 2D arrays (i.e. also on images) it is common to use the external module NumPy. This is the basis for all further mentioned modules supporting image processing and display. There are at least 3 main packages supporting image processing:

- a pair of modules: the *ndimage* module from the *SciPy* library – contains functions for processing images (including multidimensional images), and the *pyplot* module from the *Matplotlib* library – for displaying images and plots,
- *PILLOW* module (part of the non-expanded PIL module)
- *cv2* module is an frontend to the popular *OpenCV* library.

In our course we will rely on *OpenCV*, although *Matplotlib* and occasionally *ndimage* will also be used – mainly in situations where the functions in *OpenCV* do not have adequate functionality or are less convenient to use (e.g. displaying images).

1.3 Input/output operations

1.3.1 Reading, displaying and saving an image using OpenCV

Exercise 1.1 Perform a task in which you practice handling files using OpenCV.

1. From the course page, download the image *mandril.jpg* and place it in your own working directory.
2. Run the program – *spyder*, *pycharm*, *vscode* – from the console or using the icon. Create a new file and save it in your own working directory.
3. In the file, load the module `cv2` (`import cv2`). Test loading and displaying images using this library – use the function `imread` for loading, example usage: `I = cv2.imread('mandril.jpg')`
4. Displaying a picture requires at least two functions – `imshow()` creates a window and starts the display, and `waitKey()` shows the picture for a given period of time (argument 0 means wait for a key to be pressed). When the program is finished, the window is automatically closed, but on condition that it is terminated by pressing any key.

Attention! Closing the window via the button on the window bar will loop the program and you will have to abort it:

- VSCode – stop the program in the terminal by using the keyboard shortcut 'CTRL+Z' or closing the terminal.

R For especially high-resolution images, it is useful to use the function `namedWindow()` with the same name as in the function `imshow()`. We declare it before the `imshow()` function.

5. An unnecessary window can be closed using `destroyWindow` with the appropriate parameter, or all open windows can be closed using `destroyAllWindows`.

```
I = cv2.imread('mandril.jpg')
cv2.imshow("Mandril",I)          # display
cv2.waitKey(0)                  # wait for key
cv2.destroyAllWindows()         # close all windows
```

R The window does not necessarily need to be displayed on top. It is good practice to use the `cv2.destroyAllWindows()` function.

6. Saving to a file is performed by the function `imwrite` (please note the format change from *jpg* to *png*):

```
cv2.imwrite("m.png",I) # zapis obrazu do pliku
```

7. The image displayed is in colour, which means that it consists of 3 channels. In other words, it is represented as a 3D array. There is often a need to view the value of this array. This can be done in one-dimensional form, but is better done with a two-dimensional array.

- VSCode – it is necessary to run the program in *Debug* mode and set the breakpoint. Under *Run and Debug* and in the *Variables* section you will find the variable *I*. To view the values of the *I* matrix, select *View Value in Data Viewer* under the right mouse button. In the *Watch* section it is possible to refer to a specific element of the *I* array or perform appropriate operations on it.

In all cases a 2D array is displayed which is a slice of the 3D array, however by default

this slice is in the wrong axis – a single line in 3 components is displayed. To get the whole image displayed for one component, change the axis.

- VSCode – under *SLICING* section, select the appropriate axis – set to 2 or select the appropriate indexing and press the *Apply* button.

Other components are available:

- VSCode – *Index* field.

8. Access to image parameters can be useful at work:

```
print(I.shape) # dimensions /rows, columns, depth/
print(I.size)  # number of bytes
print(I.dtype) # data type
```

The print function is a display to the console.



1.3.2 Loading, displaying and saving an image using the *Matplotlib* module

An alternative way to implement input/output operations is to use the *Matplotlib* library. Then the handling of load/display etc. is similar to that known from the Matlab package. Documentation of the library is available online [Matplotlib](#).

Exercise 1.2 Do a task in which you practice handling files using the *Matplotlib* library. To keep some order in your code, create a new source code file.

1. Load *pyplot* module.

```
import matplotlib.pyplot as plt
```

(as allows you to shorten the name to be used in the project)

2. Load the same image *mandril.jpg*

```
I = plt.imread('mandril.jpg')
```

3. The display is implemented very similarly to the Matlab package:

```
plt.figure(1)      # create figure
plt.imshow(I)      # add image
plt.title('Mandril') # add title
plt.axis('off')     # disable display of the coordinate system
plt.show()         # display
```

4. Image storage:

```
plt.imsave('mandril.png', I)
```

5. During the lab, it can also be useful to display certain elements in the image – such as points or frames.

6. Adding the display of points:

```
x = [ 100, 150, 200, 250]
y = [ 50, 100, 150, 200]
plt.plot(x,y,'r.',markersize=10)
```

Attention! Commas are necessary (unlike in Matlab) when setting array values. In the plot command, the syntax is similar to Matlab – 'r' – colour, '.' – dot, and the size of the marker.

Full list of possibilities in the documentation.

7. Adding rectangle display – drawing shapes i.e. rectangles, ellipses, circles etc. is available in *matplotlib.patches*. Please note the comments in the code below.

```
from matplotlib.patches import Rectangle # add at the top of the file

fig, ax = plt.subplots(1) # instead of plt.figure(1)

rect = Rectangle((50,50),50,100,fill=False, ec='r'); # ec - edge colour
ax.add_patch(rect) # display
plt.show()
```

1.4 Colour space conversion

1.4.1 OpenCV

The function *cvtColor* is used to convert the colour space.

```
IG = cv2.cvtColor(I, cv2.COLOR_BGR2GRAY)
IHSV = cv2.cvtColor(I, cv2.COLOR_BGR2HSV)
```

R Please note that in OpenCV the reading is in **BGR** order, not RGB. This can be important when manually manipulating pixels. For a full list of available conversions with the relevant formulas in [OpenCV documentation](#)

Exercise 1.3 Convert the colour space of the given image.

1. Convert the *mandril.jpg* image to grayscale and HSV space. Display the result.
2. Display the H, S, V components of the image after conversion.

R Please note that in contrast to e.g. the Matlab package, here the indexing is from 0.

Useful syntax:

```
IH = IHSV[:, :, 0]
IS = IHSV[:, :, 1]
IV = IHSV[:, :, 2]
```

1.4.2 Matplotlib

Here the choice of available conversions is quite limited. Therefore, we need to use the formulas for colour space conversion.

1. RGB to greyscale.

$$G = 0.299 \cdot R + 0.587 \cdot G + 0.144 \cdot B \quad (1.1)$$

The formula can be represented as a function:

```
def rgb2gray(I):
    return 0.299*I[:, :, 0] + 0.587*I[:, :, 1] + 0.114*I[:, :, 2]
```

R The colour map must be set when displaying. Otherwise the image will be displayed in the default, which is not grayscale: `plt.gray()`

2. RGB to HSV.

```
import matplotlib # add at the top of the file
I_HSV = matplotlib.colors.rgb_to_hsv(I)
```

1.5 Scaling, rescaling with OpenCV

Exercise 1.4 Scale the image with *mandrill*. Use the *resize* function to scale.

Example of use:

```
height, width = I.shape[:2] # retrieving elements 1 and 2, i.e. the corresponding
                             height and width
scale = 1.75 # scale factor
Ix2 = cv2.resize(I, (int(scale*height), int(scale*width)))
cv2.imshow("Big Mandrill", Ix2)
```

1.6 Arithmetic operations: addition, subtraction, multiplication, difference module

The images are matrices, so the arithmetic operations are quite simple – just like in the Matlab package. Of course, remember to convert to the appropriate data type. Usually double format will be a good choice.

Exercise 1.5 Perform arithmetic operations on the image *lena*.

1. Download the image *lena* from the course page, then load it using a function from OpenCV – add this code snippet to the file that contains the image loader *mandril*. Perform the conversion to grayscale. Add the matrices containing Mandril and Leny in grayscale. Display the result.
2. Similarly perform the subtraction and multiplication of images.
3. Implement a linear combination of images.
4. An important operation is the module from image difference. It can be done manually – conversion to the appropriate type, subtraction, module (`abs`), conversion to `uint8`. An alternative is to use the function *absdiff* from OpenCV. Please calculate the modulus from the difference of the mandril and Lena grayscale image.

R To display the image correctly, the data type must be changed to uint type. Appropriate conversion:

```
import numpy as np
cv2.imshow("C", np.uint8(C))
```

1.7 Histogram calculation

Calculating the histogram can be done using the function `textcalcHist`. But before we get to that, let's remind ourselves of the basic control structures in Python – functions and subroutines. Please complete the following function yourself, which calculates a histogram from an image with 256 grayscale values:

```
def hist(img):
    h=np.zeros((256,1), np.float32) # creates and zeros single-column arrays
    height, width =img.shape[:2] # shape - we take the first 2 values
    for y in range(height):
        ...
    return h
```

Attention! In Python, indentation is important as it determines the block of a function, or loop!

The histogram can be displayed using the function `plt.hist` or `plt.plot` from the *Matplotlib* library.

The function `calcHist` can count the histogram of several images (or components). However, the form of the formula is most commonly used:

```
hist = cv2.calcHist([IG],[0],None,[256],[0,256])
# [IG] -- input image
# [0] -- for greyscale images there is only one channel
# None -- mask (you can count the histogram of a selected part of the image)
# [256] -- number of histogram bins
# [0 256] -- the range over which the histogram is calculated
```

Please check that the histograms obtained by both methods are the same.

1.8 Histogram equalisation

Histogram equalisation is a popular and important pre-processing operation.

1. Classical histogram equalization is a global method, it is performed on the whole image. There is a ready-to-use function for this type of equalization in the OpenCV library:

```
IGE = cv2.equalizeHist(IG)
```

2. CLAHE Histogram Equalization (*Contrast Limited Adaptive Histogram Equalization*) – is an adaptive method that improves lighting conditions in an image by using histogram equalization for individual parts of the image rather than the whole image. The method works as follows:

- division of the image into disjoint (square) blocks,
- calculation of the histogram in blocks,
- performing a histogram equalisation, where the maximum height of the histogram is limited and the excess is redistributed to adjacent intervals,
- interpolation of pixel values based on calculated histograms for given blocks (four neighbouring quadrant centres are taken into account).

For details, see [Wiki](#) and [tutorial](#).

```
clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(8,8))
# clipLimit - maximum height of the histogram bar - values above are distributed
# among neighbours
# tileGridSize - size of a single image block (local method, operates on separate
# image blocks)
```

```
I_CLAHE = clahe.apply(IG)
```

Exercise 1.6 Run and compare both equalisation methods. ■

1.9 Filtration

Filtering is a very important group of operations on images. As an exercise, please run:

- Gaussian filtration (GaussianBlur)
- Sobel filtration (Sobel)
- Laplasian filter (Laplacian)
- Median filtration (medianBlur)

 [OpenCV documentation – filtration.](#)

Please also note the other functions available:

- bilateral filtration,
- Gabor filters,
- morphological operations.

Exercise 1.7 Run and compare both equalisation methods. ■



Laboratory 2

2	Detection of foreground moving objects
	17
2.1	Image sequence loading
2.2	Frame subtraction and Binarization
2.3	Morphological operations
2.4	Labeling and simple analysis
2.5	Evaluation of foreground object detection results
2.6	Example results of the algorithm for foreground moving object detection

2. Detection of foreground moving objects

What are foreground objects ?

Those that are of interest to us (for the application under consideration). Typically: people, animals, cars (or other vehicles), luggage (potential bombs). So the definition is closely related to the target application.

Is foreground object segmentation a moving object segmentation ?

No. Firstly, an object that has stopped may still be of interest to us (e.g. a man standing in front of a pedestrian crossing). Second, there are a whole range of moving scene elements that are not of interest to us. Examples include flowing water, a fountain, moving trees and bushes, etc. Note also that usually moving objects are taken into account during image processing. So foreground object detection can and should be supported by moving object detection.

The simplest method to detect moving objects is to perform subtraction of consecutive (adjacent) frames. In this exercise we will realize a simple subtraction of two frames, combined with binarization, connected-component labeling and analysis of the resulting objects. Finally, we will try to consider the subtraction result as a result of foreground object segmentation and check the quality of this segmentation.

2.1 Image sequence loading

Exercise 2.1 Image sequence loading

1. Download the appropriate test sequence from the *UPeL* platform. In the exercise, we will focus on the sequence *pedestrians*. The sequences used come from a dataset from the changedetection.net page. The dataset contains labelled sequences, i.e. each image frame has an object mask - a reference mask (*ground truth*). On these, each pixel has been assigned to one of five categories: background (0), shadow (50), beyond the area of

interest (85), unknown motion (170) and foreground objects (255) – the corresponding greyscale levels are given in brackets.

For the exercise, we will only be interested in the split between foreground objects and the other categories. In addition, the folder contains a region of interest (ROI) mask and a text file with the time interval for which the results should be analysed (temporalROI.txt) – details follow later in the exercise.

2. The following example code allows the sequence to be loaded:

```
for i in range(300,1100):
    I = cv2.imread('input/in%06d.jpg' % i)
    cv2.imshow("I",I)
    cv2.waitKey(10)
```

R We assume that the sequence was extracted in the same folder as the source file (otherwise you need to add the appropriate path).

You must use the `waitKey` function. Otherwise the displayed image will not be refreshed.

3. Add to the loop the option to analyse every *i*-th frame – the *range* function used may have as a third parameter: *step*. Experiment with its value now (when displaying the video) and later (when detecting moving objects).

2.2 Frame subtraction and Binarization

In order to detect the moving element we subtract two consecutive frames. It should be noted that we are concerned exactly with calculating the modulus from the difference.

R To avoid problems with unsigned number subtraction (*uint8*), perform a conversion to type *int* – `IG = IG.astype('int')`.

For simplicity of consideration, it is better to convert to greyscale in advance. To begin with, you need to handle the first iteration in some way. For example – read the first frame before the loop and consider it as the previous one. Later, at the end of the loop, add the assignment `previous frame = current frame`. Test display the subtraction results – you should see mostly edges.

Binarisation can be performed using the following syntax:

```
B = 1*(D > 10)
```

R In that case, `1*` means converting from a logical type to a numeric type. The correct display is a separate issue – you need to change the range (multiply by 255) and convert to *uint8* (you may need to import the *numpy* library).

The threshold should be chosen so that objects are relatively visible. Please pay attention to compression artifacts.

An alternative is to use a function built into OpenCV:

```
(T, thresh) = cv2.threshold(D,10,255,cv2.THRESH_BINARY)
# D -- input array
# 10 -- threshold value
```

```
# 255 -- maximum value to use with the THRESH_BINARY and THRESH_BINARY_INV
        thresholding types
# cv2.THRESH_BINARY -- thresholding type
# T - our threshold value
# thresh - output image
```

- R** The first argument of the returned values by the `cv2.threshold` function is the binarization threshold (this is useful when using automatic threshold determination, e.g. with the Otsu or triangle method). See the OpenCV documentation for details.

2.3 Morphological operations

Exercise 2.2 The resulting image is quite noisy. Please perform filtering using erosion and dilation (`erode` and `dilate` from OpenCV).

- R** The aim of this stage is to obtain a maximally visible silhouette, with minimum distortions. To improve the effect it is worth adding a median filtering step (before morphology) and possibly correcting the binarisation threshold.

2.4 Labeling and simple analysis

In the next step, we will perform a filtering of the obtained result. We will use indexing (assigning labels to groups of connected pixels) and calculating the parameters of these groups. The function `connectedComponentsWithStats` is used for this. Function call:

```
retval, labels, stats, centroids = cv2.connectedComponentsWithStats(B)
# retval -- total number of unique labels
# labels -- destination labeled image
# stats -- statistics output for each label, including the background label.
# centroids -- centroid output for each label, including the background label.
```

- R** When displaying the *labels* image, you need to set the format properly and add scaling. You can use information about the number of objects found for scaling.

```
cv2.imshow("Labels", np.uint8(labels / retval * 255))
```

Then display the surrounding rectangle, field and index for the largest object. Below is a sample solution to the task. Please run it and possibly try to optimise it.


```
I_VIS = I # copy of the input image

if (stats.shape[0] > 1): # are there any objects

    tab = stats[1:,4] # 4 columns without first element
    pi = np.argmax( tab )# finding the index of the largest item
    pi = pi + 1 # increment because we want the index in stats, not in tab
    # drawing a bbox
    cv2.rectangle(I_VIS, (stats[pi,0], stats[pi,1]), (stats[pi,0]+stats[pi,2], stats[pi,1]+stats[pi,3]), (255,0,0), 2)
    # print information about the field and the number of the largest element
    cv2.putText(I_VIS, "%f" % stats[pi,4], (stats[pi,0], stats[pi,1]), cv2.FONT_HERSHEY_SIMPLEX, 0.5, (255,0,0))
```

```
cv2.putText(I_VIS, "%d" % pi, (np.int(centroids[pi,0]), np.int(centroids[pi,1])), cv2.FONT_HERSHEY_SIMPLEX, 1, (255,0,0))
```

Comments on the sample solution:

- `stats.shape[0]` is the number of objects. Since the function also counts the object with index 0 (i.e. background), in the conditional function check if there are objects the condition is `> 1`.
- the next two lines are the calculation of the maximum index from column number 4 (field).
- the resulting index should be incremented, as we have omitted element 0 (background) from the analysis.
- We use the `rectangle` function from OpenCV to draw a surrounding rectangle in the image. The syntax `"%f" % stats[pi,4]` allows us to output the value in the appropriate format (f - float). The next parameters are the coordinates of the two opposite vertices of the rectangle. Then the colour in format (B,G,R) and finally the line thickness. See the function documentation for details.
- The function `putText` is used to output text on the image. The coordinates of the lower left vertex are given to determine the position of the text box. The next parameters are the font (full list in the documentation), size and colour.
-  `I_VIS` is the input image before conversion to greyscale.

Exercise 2.3 Use the function `cv2.connectedComponentsWithStats` to label and calculate the parameters of the resulting objects. Display the surrounding rectangle, field and index for the largest object. Based on the tips and examples in the text above, try to optimize the solution.

2.5 Evaluation of foreground object detection results

In order to evaluate the foreground object detection algorithm, and preferably in a relatively objective manner, the results returned by it, i.e. the object mask, should be compared with the reference mask (*groundtruth*). The comparison takes place at the level of individual pixels. If shadows are excluded (which is what we assumed at the beginning), four situations are possible:

- *TP – true positive* – a pixel belonging to a foreground object is detected as a pixel belonging to a foreground object,
- *TN – true negative* – a pixel belonging to the background is detected as a pixel belonging to the background,
- *FP – false positive* – a pixel belonging to the background is detected as a pixel belonging to a foreground object,
- *FN – false negative* – a pixel belonging to an object is detected as a pixel belonging to the background.

A series of metrics can be counted from the listed coefficients. We will use three: *precision* - *P*, *recall* - *R* and *F1* score. These are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (2.1)$$

$$R = \frac{TP}{TP + FN} \quad (2.2)$$

$$F1 = \frac{2PR}{P + R} \quad (2.3)$$

The F1 score is in the range $[0; 1]$, with the higher its value, the better it is.

Exercise 2.4 Implement the computation of the metrics P , R and $F1$.

1. In the first step, define the global counters TP , TN , FP , FN , and calculate the desired values of P , R and $F1$ when the loop finished.
2. Add to the loop the loading of a reference mask – analogous to an image. It is available in the *groundtruth* folder.
3. The next step is to compare the object mask and the reference mask. The simplest solution is for loop over the whole image, in each iteration we check the pixel similarity. Of course, this is not a very efficient approach. You can try using Python's capabilities in implementing matrix operations. Create appropriate logical conditions, for example:

```
TP_M = np.logical_and((B == 255), (GTB == 255)) # logical product of the
matrix elements
TP_S = np.sum(TP_M) # sum of the elements in the matrix
TP = TP + TP_S # update of the global indicator
```

However, we only perform the calculation if a valid reference map is available. To do this, check the dependence of the frame counter and the value from the *temporalROI.txt* file – it must be within the range described there. Example code that loads the range (by the way, please note how text files are handled):

```
f = open('temporalROI.txt', 'r') # open file
line = f.readline() # read line
roi_start, roi_end = line.split() # split line
roi_start = int(roi_start) # conversion to int
roi_end = int(roi_end) # conversion to int
```

4. Run the calculations for the consecutive frame subtraction method. Note the value of $F1$.
5. Perform calculations for the other two sequences – *highway* and *office*.

R Information on the following classes.

1. In further exercises we will learn more algorithms and functionalities of the Python language. However, we do not put special emphasis on learning the Python language, but on using in practice the successive algorithms appearing in the lectures. The subject Advanced Vision Systems is not a Python course!
2. The solutions presented should always be considered as examples. The problem can certainly be solved differently, and sometimes better.
3. In the next exercises the level of detail of the solution description will decrease. We therefore encourage you to actively use the documentation for Python, OpenCV and materials/tutorials found on the Internet :).

2.6 Example results of the algorithm for foreground moving object detection

In order to better verify the different steps of the proposed method for foreground moving object detection, an example result for an image with index 350 will be presented.



Figure 2.1: An example of the result of each stage of the algorithm. From left: input image with bounding box, image after binarisation, image after median filtering and morphological operations (erode, dilate), image representing labels after labelling, reference image – ground truth.



Laboratory 3

3	Foreground object segmentation	25
3.1	Main objectives	
3.2	Theory of foreground object segmentation	
3.3	Methods based on frame buffer	
3.4	Approximation of mean and median (so-called sigma-delta)	
3.5	Updating policy	
3.6	OpenCV – GMM/MOG	
3.7	OpenCV – KNN	
3.8	Example results of the algorithm for foreground object segmentation	
3.9	Using a neural network to segment foreground objects.	
3.10	Additional exercises	

3. Foreground object segmentation

3.1 Main objectives

- You will become familiar with the issue of foreground object segmentation and the problems associated with it.
- You will be able to implement simple background modelling algorithms based on a frame buffer – analyzing their pros and cons.
- You will be able to implement the running average and approximate median method.
- You will become familiar with the methods available in *OpenCV* – Gaussian Mixture Model (also called Mixture of Gaussians) and a method based on the KNN (K-Nearest Neighbours) algorithm,
- You will become familiar with neural network architecture and an example application.


3.2 Theory of foreground object segmentation

What is foreground object segmentation?

Segmentation is the extraction of a certain group of objects present in an image. It should be noted that it is not necessarily the same as classification, where the output is the assignment of an object to a specific class: e.g. a car (or even a type), a pedestrian, etc. The definition of a foreground object is not very strict – they are all objects relevant to the application. For example for video surveillance: people, cars and maybe animals.

What is the background model?

In the simplest case, it is an image of an empty scene, i.e. without objects of interest.

 In more advanced algorithms, the background model has no explicit form (it is not an image) and cannot simply be displayed (examples of such methods: GMM, ViBE, PBAS).

Initialisation of the background model

When the algorithm is started (also restarted), it is necessary to perform an initialization of the background model, i.e., to determine the values of all parameters (variables) that represent this model. In the simplest case, the first frame of the analysed sequence (the first N frames in the case of methods with buffer) is taken as the initial background model. This approach works well under the assumption that there are no foreground objects on that particular frame (sequence of frames). Otherwise, the initial model will contain errors that will be more or less difficult to eliminate in further operation - it depends on the specific algorithm. In the literature this issue is referred to as *bootstrap*.

More advanced initialisation methods rely on temporal and even spatial pixel analysis for a certain initial sequence. The assumption here is that the background is visible most of the time and is more spatially consistent than the objects.

Background modelling, background generation

The question is whether it is not enough to take an image of an empty scene, save it and use it throughout the algorithm? In the general case no, but let's start with a special case. If our system is monitoring a room where the lighting is strictly controlled and any previously unforeseen and approved changes should not take place (e.g. a forbidden zone inside a power plant, a bank vault, etc.), then the simplest approach with a static background will work.

Problems arise when the lighting of a scene changes due to a change in the time of day or when additional lights are switched on, etc. Then the actual background will change and our static background model will not be able to adapt to this change. The second, more difficult case is a change in the position of scene elements: e.g. someone moves a bench/chair/flower. This change should not be detected, or at least after a fairly short time it should be taken into account in the background model. Otherwise it will cause the generation of false detections (known as *ghost*) and consequently the generation of false alarms.

The conclusion is as follows. A good background model should be characterised by its ability to adapt to changes in the scene. The phenomenon of model updating is referred to in the literature as background generation or background modelling.

3.3 Methods based on frame buffer

The exercise will present two methods: the buffer mean and the buffer median. The buffer size will be assumed to be $N = 60$ samples. In each iteration of the algorithm it is necessary to remove the last frame from the buffer, add the current one (the buffer works on the principle of FIFO queue) and calculate the mean or median from the buffer.

1. First, declare a buffer of size $N \times YY \times XX$ (`np.zeros` function), where YY and XX are the height and width of the frame from the *pedestrians* sequence.

```
BUF = np.zeros((YY, XX, N), np.uint8)
```

You should also initialize the counter for the buffer (let it be called e.g. `iN`) with the value 0. The parameter `iN` is like a pointer that points to a place in the buffer from which the last frame should be removed and the current frame assigned.

2. The buffer handling should be as follows. In the loop under the variable `iN`, store the current frame in greyscale.

```
BUF[:, :, iN] = IG;
```

Then increment the counter iN and check if it does not reach the buffer size (N). If it does, it needs to be set to 0.

3. The computation of the mean or median is implemented by the functions `mean` or `median` from the *numpy* library. As a parameter of the function you specify the dimension for which the value is to be calculated (remember to index from zero). Then convert the result to `uint8`.
4. Finally, we perform background subtraction, i.e. we subtract the model from the current scene and binarise the result – analogous to the difference of adjacent frames. We also use median filtering of object masks and/or morphological operations.
5. Using the code created in the previous exercise, compare the performance of the method with the mean and median. Note the values of the $F1$ indicator for both cases. Calculating the median can take a long time. Consider why the median works better. Check the performance on other sequences – in particular for the sequence *office*. Observe the phenomenon of blurring and silhouette blending into the background and the formation of *ghost*.

Exercise 3.1 Implement the algorithms (median and mean) based on the above description. ■

3.4 Approximation of mean and median (so-called sigma-delta)

Using a sample buffer is resource-intensive. There are methods for faster calculation of mean, but in case of median such methods do not exist. Therefore, methods that do not require a buffer are more likely to be used. In the case of the mean approximation the relation is used:

$$BG_N = \alpha I_N + (1 - \alpha) BG_{N-1} \quad (3.1)$$

where: I_N – current frame, BG_N – background model, α – weight parameter, typically 0.01 – 0.05, although the value depends on the specific application.

The approximation of the median is obtained using the relation:

$$\begin{aligned} \text{if } BG_{N-1} < I_N \text{ then } BG_N &= BG_{N-1} + 1 \\ \text{if } BG_{N-1} > I_N \text{ then } BG_N &= BG_{N-1} - 1 \\ \text{otherwise } BG_N &= BG_{N-1} \end{aligned} \quad (3.2)$$

Exercise 3.2 Implement both methods.

1. Take the first frame of the sequence as the first background model. Note the value of the $F1$ indicator for the two methods (set the *alpha* parameter to 0.01). Observe the running time.
2. Experiment with the value of the *alpha* parameter. See how the parameter value affects the background model.


For the running average method it is very important that the background model is floating point (*float64*).

For the pattern 3.1 the avoidance of using a loop over the whole image is obvious. For the dependency 3.2 this is also possible. It is worthwhile to use the documentation of *numpy*. Additionally, please note that in Python it is possible to perform arithmetic operations on Boolean values – `False == 0` while `True == 1`.

3.5 Updating policy

So far we have followed a liberal update policy – we have updated everything. We will now try to use a conservative approach.

Exercise 3.3 1. For the selected method, implement a conservative update approach. Check its operation.

 Simply remember the previous object mask and use it appropriately in the update procedure.

2. Note the value of the F1 indicator and the background model. Are there any errors in it? ■

3.6 OpenCV – GMM/MOG

One of the most popular foreground object segmentation methods is available in the OpenCV library: Gaussian Mixture Models (GMM) or Mixture of Gaussians (MoG) - both names appear in parallel in the literature. In a most brief way: a scene is modelled by several Gaussian distributions (mean of brightness (colour) and standard deviation). Each distribution is also described by a weight parameter that captures how often it was used (observed in the scene – probability of occurrence). The distributions with the largest weight parameters represent the background. Segmentation is based on calculating the distance between the current pixel and each of the distributions. Then if the pixel is similar to the distribution considered as background, it is classified as background, otherwise it is considered as object. The background model is updated in a way similar to the equation 3.1. If you are interested, we refer you to the literature, e.g. [article](#).

Exercise 3.4 This method is an example of a multivariant algorithm – the background model has several possible representations (variants).

1. Open the documentation for [OpenCV](#). Go to *Video Analysis->Motion Analysis*. We are interested in the class `BackgroundSubtractorMOG2`. To create an object of this class in Python3 use `createBackgroundSubtractorMOG2`. Using this object in a loop for each image we use its method `apply`.
2. To begin, we will run the code with the default values.
3. Please experiment with the parameters: *history*, *varThreshold* and disable shadow detection. It is also possible to manually set the learning rate in the `apply()` method – *learningRate*. There is a GMM version available in the OpenCV package which is slightly different from the original – including a shadow detection module and a dynamically changing number of Gaussian distributions. Article references are available in the OpenCV documentation.
4. Determine the F1 parameter (for the method without shadow detection). Observe how the method works. Note that the background model cannot be displayed. ■

3.7 OpenCV – KNN

The second method available in OpenCV is KNN algorithm (`BackgroundSubtractorKNN`).

Exercise 3.5 Please run the KNN method and evaluate it. ■

3.8 Example results of the algorithm for foreground object segmentation

In order to better verify the individual steps of the proposed methods for foreground object segmentation, example results for selected image frames will be presented.

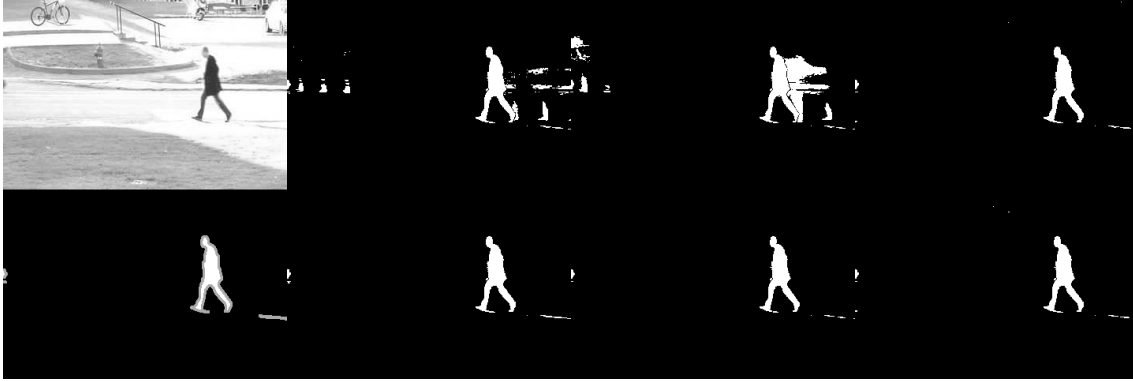


Figure 3.1: Example of the result of the different versions of the algorithm for foreground object segmentation for the *pedestrians* dataset. The frame index is 600. The first column from the left is the input image together with the reference image. The first row is for algorithms using the mean, the second row uses the median. Sequentially from the second column the result of the algorithm is presented: liberal update policy, then the function approximation with a liberal update policy and the function approximation with a conservative update policy.



Figure 3.2: Example of the result of the different versions of the algorithm for foreground object segmentation for the *highway* dataset. The frame index is 1200. The first column from the left is the input image together with the reference image. The first row is for algorithms using the mean, the second row uses the median. Sequentially from the second column the result of the algorithm is presented: liberal update policy, then the function approximation with a liberal update policy and the function approximation with a conservative update policy.

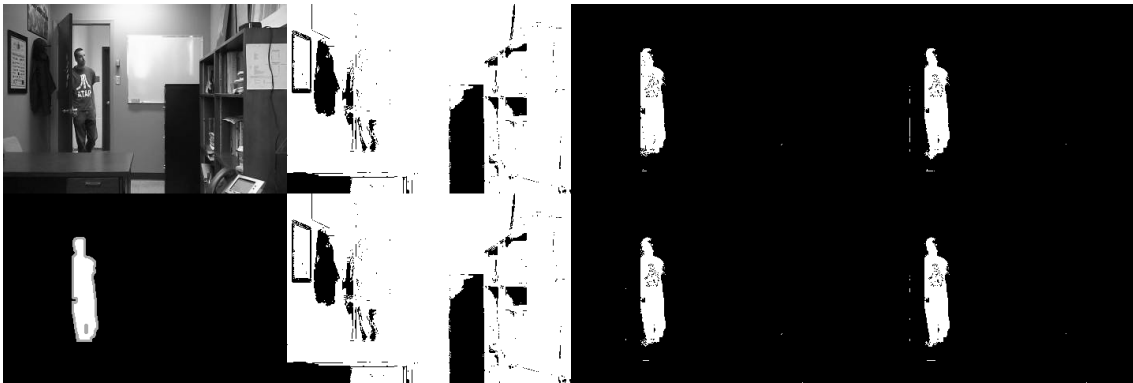


Figure 3.3: Example of the result of the different versions of the algorithm for foreground object segmentation for the *office* dataset. The frame index is 600. The first column from the left is the input image together with the reference image. The first row is for algorithms using the mean, the second row uses the median. Sequentially from the second column the result of the algorithm is presented: liberal update policy, then the function approximation with a liberal update policy and the function approximation with a conservative update policy.

3.9 Using a neural network to segment foreground objects.

Neural network architecture-based methods can also be used to segment foreground objects. An example of this is the work ¹. The proposed approach is to use a convolutional neural network called BSUV-Net (Background Subtraction Unseen Videos Net). The architecture is based on a U-net structure with residual connections, which is divided into an encoder and a decoder. The input to network consists of the current frame and two background frames captured at different time scales along with their semantic segmentation maps. The output is a mask containing only foreground objects.

You can find details about this neural network in the article <https://arxiv.org/abs/1907.11371>. The architecture with sample data and trained network models is made available on Github <https://github.com/ozantezcan/BSUV-Net-inference>.

Exercise 3.6 Run the BSUV-Net convolutional neural network.

- Download all the files for the neural network from the UPeL platform.
- The pedestrians database will be used, but appropriately scaled to the size of the network input and converted to video sequences,
- Open `infer_config_manualBG.py` file and modify the paths to:
 - in class `SemanticSegmentation` specify the absolute path to the segmentation folder – variable `root_path`,
 - in class `BSUVNet` specify the path to the network model `BSUV-Net-2.0.mdl` in the `trained_models` folder – variable `model_path`,
 - in class `BSUVNet` specify the path to an image that contains only the background – the first image in the set `pedestrians` – variable `empty_bg_path`
- In the `inference.py` file specify the path to the network input, i.e. the pedestrians video sequence – variable `inp_path` and a path to where the network output is to be stored (path with the file name and file extension) – variable `out_path`.

3.10 Additional exercises

3.10.1 Initialization of the background model in the presence of foreground objects on the scene

Exercise 3.7 Additional exercise 1

If we take a method in which we have used a conservative update approach and start the sequence analysis not with frame 1 but with 1000 then we will see in practice the disadvantages of assuming that the first frame in the sequence is the initial background model (*ghosty* etc.). Propose a solution to the problem. In your application, demonstrate the advantage of your own approach over initialization based on the first frame.

Figure 3.4 shows an example of a solution to the problem.

- R** It seems a good idea to buffer, for example, 30 or more frames and perform a frequency analysis of the occurrence of each shade of brightness in the sequence – independently for each pixel. The background should be considered to be that which was observed most frequently. Other effective ideas are also welcome.

¹Tezcan, Ozan and Ishwar, Prakash and Konrad, Janusz; BSUV-Net: A Fully-Convolutional Neural Network for Background Subtraction of Unseen Videos, <https://arxiv.org/abs/1907.11371>

A different sequence can be used in the task.

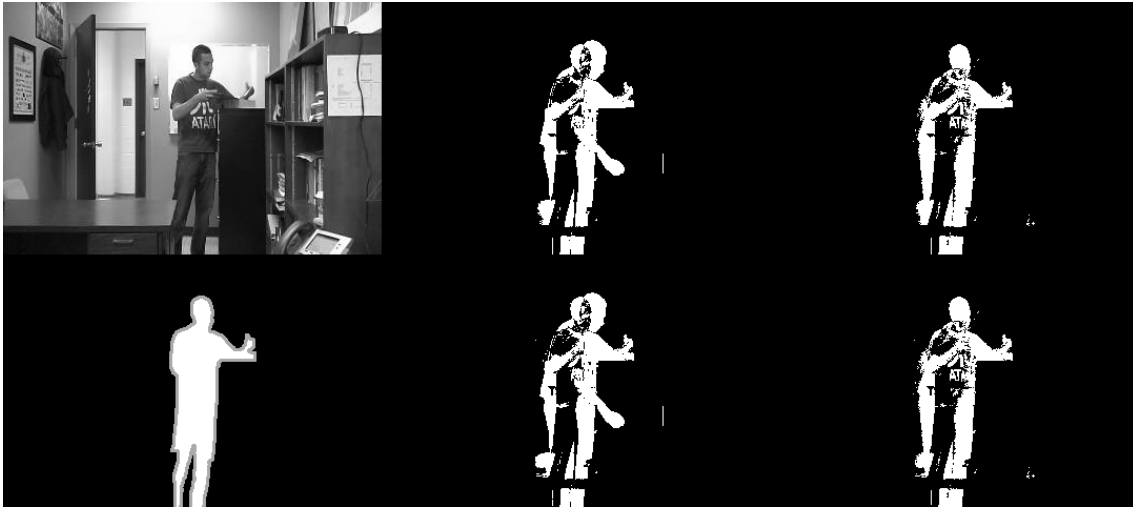


Figure 3.4: Example result of the proposed solution for *office* dataset. The frame index is 800. The first column from the left is the input image together with the reference image. The first row is for algorithms using the mean, the second row uses the median. Sequentially from the second column the result of the algorithm is presented: function approximation with conservative update policy without initialization buffer, function approximation with conservative update policy with initialization buffer.

3.10.2 Implementation of ViBE and PBAS methods

Exercise 3.8 Additional exercise 2

There are articles on the course website which describe the ViBE and PBAS methods. The task is to implement and evaluate them. Please follow the order, as PBAS is an extension of ViBE.

Figure 3.5 shows an example solution to the problem.



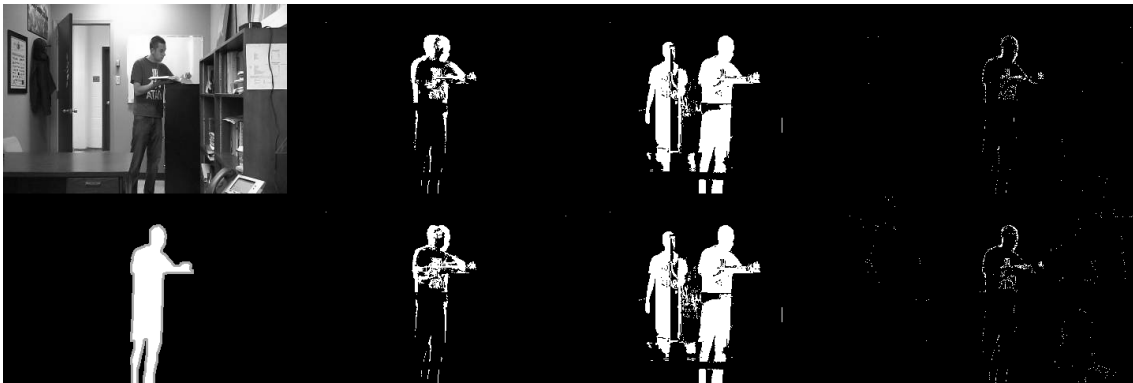


Figure 3.5: Example result of the proposed solution for *office* dataset. The frame index is 1000. The first column from the left is the input image together with the reference image. Then the approximation of the mean and median with a liberal update policy, the approximation of the mean and median with a conservative update policy and the last column is the ViBe algorithm (top) and the PBAS algorithm (bottom).