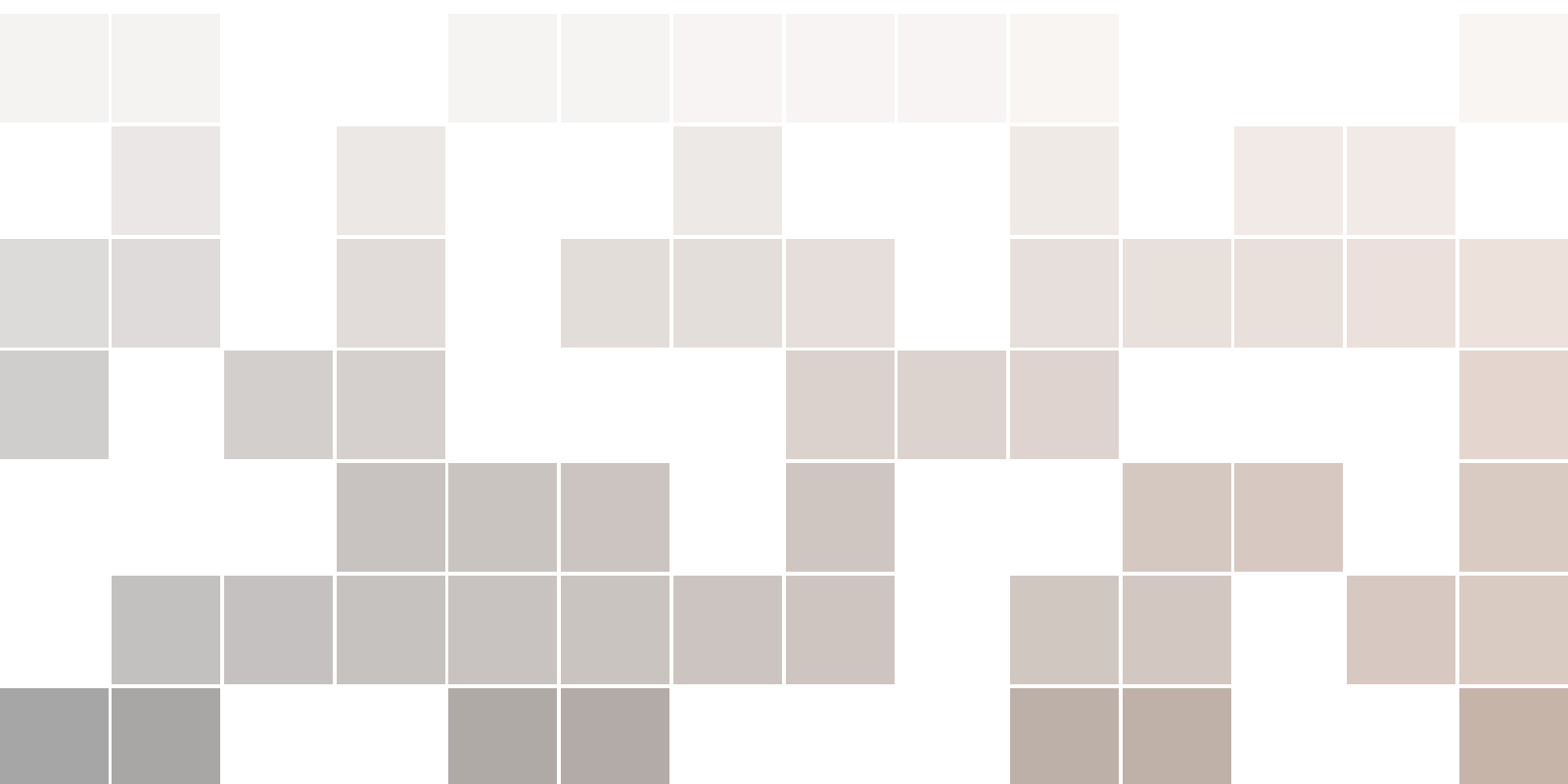




# Zaawansowane algorytmy wizyjne

Skrypt do ćwiczeń laboratoryjnych

Piotr Pawlik, Tomasz Kryjak, Mateusz Wąsala



Copyright © 2023

Piotr Pawlik

Tomasz Kryjak

Mateusz Wąsala

ZESPÓŁ WBUDOWANYCH SYSTEMÓW WIZYJNYCH

LABORATORIUM SYSTEMÓW WIZYJNYCH

WEAiIB, AGH w KRAKOWIE

*Wydanie trzecie, zmienione i poprawione*

*Kraków, luty 2024*

## Spis treści

I	Laboratorium 1	
1	Algorytmy wizyjne w Python 3.X – wstęp .....	7
1.1	Wykorzystywane oprogramowanie	7
1.2	Moduły Pythona wykorzystywane w przetwarzaniu obrazów	7
1.3	Operacje wejścia/wyjścia	8
1.3.1	Wczytywanie, wyświetlanie i zapisywanie obrazu z wykorzystaniem OpenCV ..	8
1.3.2	Wczytywanie, wyświetlanie i zapisywanie obrazu z wykorzystaniem modułu <i>Matplotlib</i> .....	9
1.4	Konwersje przestrzeni barw	11
1.4.1	OpenCV .....	11
1.4.2	Matplotlib .....	11
1.5	Skalowanie, zmiana rozdzielczości przy użyciu OpenCV	12
1.6	Operacje arytmetyczne: dodawanie, odejmowanie, mnożenie, moduł z różnicą	12
1.7	Wyliczenie histogramu	13
1.8	Wyrównywanie histogramu	13
1.9	Filtracja	14
II	Laboratorium 2	
2	Detekcja pierwszoplanowych obiektów ruchomych .....	17
2.1	Wczytywanie sekwencji obrazów	17

2.2	Odejmowanie ramek i binaryzacja	18
2.3	Operacje morfologiczne	19
2.4	Indeksacja i prosta analiza	19
2.5	Ewaluacja wyników detekcji obiektów pierwszoplanowych	20
2.6	Przykładowe rezultaty algorytmu do detekcji ruchomych obiektów pierwszoplanowych	22

### III

## Laboratorium 3

3	Segmentacja obiektów pierwszoplanowych .....	25
3.1	Cel zajęć	25
3.2	Segmentacja obiektów pierwszoplanowych	25
3.3	Metody oparte o bufor próbek	27
3.4	Aproksymacja średniej i mediany (tzw. sigma-delta)	28
3.5	Polityka aktualizacji	29
3.6	OpenCV – GMM/MOG	29
3.7	OpenCV – KNN	30
3.8	Przykładowe rezultaty algorytmu do segmentacji obiektów pierwszoplanowych	30
3.9	Wykorzystanie sieci neuronowej do segmentacji obiektów pierwszoplanowych	33
3.10	Zadania dodatkowe	33
3.10.1	Inicjalizacja modelu tła w warunkach obecności obiektów pierwszoplanowych na scenie .....	33
3.10.2	Implementacja metod ViBE i PBAS .....	34



# Laboratorium 1

<b>1</b>	<b>Algorytmy wizyjne w Python 3.X – wstęp 7</b>
1.1	Wykorzystywane oprogramowanie
1.2	Moduły Pythona wykorzystywane w przetwarzaniu obrazów
1.3	Operacje wejścia/wyjścia
1.4	Konwersje przestrzeni barw
1.5	Skalowanie, zmiana rozdzielczości przy użyciu OpenCV
1.6	Operacje arytmetyczne: dodawanie, odejmowanie, mnożenie, moduł z różnicy
1.7	Wyliczenie histogramu
1.8	Wyrównywanie histogramu
1.9	Filtracja



# 1. Algorytmy wizyjne w Python 3.X – wstęp

W ramach ćwiczenia zaprezentowane/przypomniane zostaną podstawowe informacje związane z wykorzystaniem języka Python do realizacji operacji przetwarzania i analizy obrazów, a także sekwencji wideo.

## 1.1 Wykorzystywane oprogramowanie

Przed rozpoczęciem ćwiczeń **proszę utworzyć** własny katalog roboczy w miejscu podanym przez Prowadzącego. Nazwa katalogu powinna być związana z Państwa nazwiskiem – zalecana forma to NazwiskoImie bez polskich znaków.

Do przeprowadzania ćwiczeń można wykorzystać jedno z trzech środowisk programowania w Pythonie – Spyder5, PyCharm lub Visual Studio Code (VSCode). Ich zaletą jest możliwość łatwego podglądu tablic dwuwymiarowych (czyli obrazów). Spyder jest prostszy, ale także ma mniejsze możliwości i więcej niedociągnięć. PyCharm jest oprogramowaniem komercyjnym udostępnianym także w wersji darmowej (wygląd zbliżony do CLion dla C/C++). W PyCharmie pracę należy zacząć od utworzenia projektu, Spyder pozwala na pracę na pojedynczych plikach (bez konieczności tworzenia projektu).

Visual Studio Code jest darmowym, lekkim, intuicyjnym i łatwym w obsłudze edytorem kodu źródłowego. Jest to uniwersalne oprogramowanie do dowolnego języka programistycznego. Konfiguracja programu do ćwiczeń sprowadza się jedynie do wybrania odpowiedniego interpretera, a dokładnie odpowiedniego środowiska wirtualnego w języku Python.

## 1.2 Moduły Pythona wykorzystywane w przetwarzaniu obrazów

Python nie posiada natywnego typu tablicowego pozwalającego na wykonanie operacji pomiędzy elementami dwóch kontenerów (w dogodny i wydajny sposób). Do operacji na tablicach 2D (czyli także na obrazach) powszechnie używa się zewnętrznego modułu *NumPy*. Jest to podstawa dla wszystkich dalej wymienionych modułów wspierających przetwarzanie i wyświetlanie obrazów. Istnieją co najmniej 3 zasadnicze pakiety wspierające przetwarzanie obrazów:

- para modułów: moduł *ndimage* z biblioteki *SciPy* – zawiera funkcje do przetwarzania obrazów (także wielowymiarowych) oraz moduł *pyplot* z biblioteki *Matplotlib* – do wyświetlania obrazów i wykresów,
- moduł *PILLOW* (fork<sup>1</sup> starszego, nierozwijanego modułu *PIL*),
- moduł *cv2* będący nakładką na popularną bibliotekę *OpenCV*.

W naszym kursie oprzemy się na *OpenCV*, aczkolwiek wykorzystywane będzie także *Matplotlib* i sporadycznie *ndimage* – głównie w sytuacjach kiedy funkcje z *OpenCV* nie mają odpowiednich funkcjonalności albo są mniej wygodne w stosowaniu (np. wyświetlanie obrazów).

## 1.3 Operacje wejścia/wyjścia

### 1.3.1 Wczytywanie, wyświetlanie i zapisywanie obrazu z wykorzystaniem OpenCV

**Ćwiczenie 1.1** Wykonaj zadanie, w którym przećwiczysz obsługę plików z wykorzystaniem *OpenCV*.

1. Ze strony kursu pobierz obraz *mandril.jpg* i umieść go we własnym katalogu roboczym.
2. Uruchom program – *spyder*, *pycharm*, *vscode* – z konsoli lub za pomocą ikony. Stwórz nowy plik i zapisz go we własnym katalogu roboczym.
3. W pliku załaduj moduł *cv2* (`import cv2`). Przetestuj wczytywanie i wyświetlanie obrazów za pomocą tej biblioteki – do wczytywania służy funkcja *imread* (przykład użycia: `I = cv2.imread('mandril.jpg')`).
4. Wyświetlenie obrazka wymaga użycia co najmniej dwóch funkcji – *imshow()* tworzy okienko i rozpoczyna wyświetlanie, a *waitKey()* pokazuje obraz przez zadany okres czasu (argument 0 oznacza czekanie na wciśnięcie klawisza). Po zakończeniu pracy okno jest automatycznie zamykane, ale pod warunkiem zakończenia przez naciśnięcie dowolnego klawisza.

**Uwaga!** Zamknięcie okna przez przycisk na belce okna spowoduje zapętlenie programu i konieczność jego przerwania:

- *Spyder* – zamknięcie konsoli *IPython*,
- *PyCharm* – wybranie pozycji 'Stop' w menu lub odpowiadającego jej przycisku,
- *VSCode* – zatrzymanie programu w terminalu poprzez użycie skrótu klawiszowego 'CTRL+Z' lub zamknięcie terminalu.

**R** W przypadku szczególnie obrazów o dużej rozdzielczości warto przed funkcją *cv2.imshow()* użyć funkcję *namedWindow()* z tą samą nazwą co w funkcji *imshow()*.

5. Niepotrzebne już okno można zamknąć za pomocą *destroyWindow* z odpowiednim parametrem lub zamknąć wszystkie otwarte okna za pomocą *destroyAllWindows*.

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

**R** Okno niekoniecznie musi zostać wyświetlone na wierzchu. Dobrą praktyką jest zawsze stosowanie funkcji *cv2.destroyAllWindows()*.

<sup>1</sup> gałąź boczna projektu



6. Zapis do pliku realizuje funkcja `imwrite` (proszę zwrócić uwagę na zmianę formatu z `jpg` na `png`):

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

7. Wyświetlany obraz jest kolorowy, co oznacza, że składa się z 3 składowych. Innymi słowy jest reprezentowany jako tablica 3D. Często istnieje potrzeba podglądu wartości tej tablicy. Można to robić w formie jednowymiarowej, ale lepiej to zrobić za pomocą macierzy dwuwymiarowej.

- Spyder – w zakładce *Variable explorer* (narzędzie podobne do *Workspace* z Matlab) wystarczy kliknąć na zmienną `I`.
- PyCharm – konieczne jest uruchomienie programu w trybie *Debug* i ustawienie breakpointa. Po zatrzymaniu na nim i kliknięciu w zakładkę *Debugger* należy kliknąć w tekst *View as Array* na końcu linii pokazującej zmienną `I`,
- VSCode – konieczne jest uruchomienie programu w trybie *Debug* i ustawienie breakpointa. W zakładce *Run and Debug* i w sekcji *Variables* znajdziemy zmienną `I`. Aby wyświetlić wartości macierzy `I` należy wybrać opcję *View Value in Data Viewer* pod prawym klawiszem myszki. W sekcji *Watch* możliwe jest odwołanie do konkretnego elementu tablicy `I` lub wykonać odpowiednie operacje na niej.

We wszystkich przypadkach wyświetlana jest macierz 2D będąca “wycinkiem” tablicy 3D, jednakże domyślnie jest to wycinek “w złej osi” – wyświetlana jest pojedyncza linia w 3 składowych. Aby uzyskać wyświetlenie całego obrazu dla jednej składowej należy zmienić oś.

- Spyder – pole *Axis* należy ustawić na 2,
- PyCharm – wycinek `I[0]` należy zmienić na `I[:, :, 0]`,
- VSCode – w sekcji *SLICING* wybrać odpowiednią oś – ustawić na 2 lub wybrać odpowiednie indeksowanie i nacisnąć przycisk *Apply*.

Pozostałe składowe są dostępne:

- Spyder – pole *Index*, wartości 1 i 2,
- PyCharm – Wycinki `I[:, :, 1]` i `I[:, :, 2]`,
- VSCode – pole *Index*.

8. W pracy przydatny może być dostęp do parametrów obrazu:

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

Funkcja `print` to oczywiście wyświetlanie na konsolę.

### **Rozliczenie ćwiczenia**

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu `.py` lub `.ipynb` w odpowiednim miejscu w zasobach kursu na UPeL.

## **1.3.2 Wczytywanie, wyświetlanie i zapisywanie obrazu z wykorzystaniem modułu *Matplotlib***

Alternatywny sposób realizacji operacji wejścia/wyjścia to użycie biblioteki *Matplotlib*. Wtedy obsługa wczytywania/wyświetlania itp. jest zbliżona do tej znanej z pakietu Matlab. Dokumentacja

biblioteki dostępna jest on-line [Matplotlib](#).

**Ćwiczenie 1.2** Wykonaj zadanie, w którym przećwiczysz obsługę plików z wykorzystaniem biblioteki Matplotlib. W celu zachowania pewnego porządku w kodzie, utwórz nowy plik z kodem źródłowym.

1. Załaduj moduł *pyplot*.

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

(as pozwala skrócić nazwę, którą trzeba będzie wykorzystywać w projekcie)

2. Wczytaj ten sam obraz *mandril.jpg*

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

3. Wyświetlanie realizuje się bardzo podobnie jak w pakiecie Matlab:

```
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. Zapisywanie obrazu:

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

5. Podczas laboratorium przydatne może być też wyświetlanie pewnych elementów na obrazku – np. punktów, czy ramek.
6. Dodanie wyświetlania punktów:

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

**Uwaga!** Przy ustalaniu wartości tablicy przecinki są niezbędne (w odróżnieniu od Matlab). W poleceniu `plot` składania podobna jak w Matlabie – 'r' – kolor, '.' – kropka oraz rozmiar markera.

Pełna lista możliwości w dokumentacji.

7. Dodanie wyświetlania prostokąta – rysowanie kształtów tj. prostokątów, elips, kół itp. jest dostępne w *matplotlib.patches*. Proszę zwrócić uwagę na komentarze w poniższym kodzie.

```
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()
```

### Rozliczenie ćwiczenia

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu .py lub .ipynb w odpowiednim miejscu w zasobach kursu na UPeL.

## 1.4 Konwersje przestrzeni barw

### 1.4.1 OpenCV

Do konwersji przestrzeni barw służy funkcja *cvtColor*.

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

**Uwaga!** Proszę zauważyć, że w OpenCV odczyt jest w kolejności BGR, a nie RGB. Może to być istotne w przypadku ręcznego manipulowania pikselami. Pełna lista dostępnych konwersji wraz ze stosownymi wzorami w [dokumentacji OpenCV](#)

**Ćwiczenie 1.3** Wykonaj konwersję przestrzeni barw podanego obrazu.

1. Dokonać konwersji obrazu *mandril.jpg* do odcieni szarości i przestrzeni HSV. Wynik wyświetlić.
2. Wyświetlić składowe H, S, V obrazu po konwersji.

#### Rozliczenie ćwiczenia

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu .py lub .ipynb w odpowiednim miejscu w zasobach kursu na UPeL.

**Uwaga!** Proszę zwrócić uwagę, że w odróżnieniu np. od pakietu Matlab, tu indeksowanie jest od 0.

Przydatna składnia:

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

### 1.4.2 Matplotlib

Tu wybór dostępnych konwersji jest dość ograniczony.

1. RGB do odcieni szarości. Można wykorzystać rozbiecie na poszczególne kanały i wzór:

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

Całość można opakować w funkcję:

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

**Uwaga!** Przy wyświetlaniu należy ustawić mapę kolorów. Inaczej obraz wyświetli się w domyślnej, która nie jest bynajmniej w odcieniach szarości: `plt.gray()`

## 2. RGB do HSV.

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

## 1.5 Skalowanie, zmiana rozdzielczości przy użyciu OpenCV

**Ćwiczenie 1.4** Przeskaluj obraz *mandril*. Do skalowania służy funkcja `resize`.

### Rozliczenie ćwiczenia

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu `.py` lub `.ipynb` w odpowiednim miejscu w zasobach kursu na UPeL. ■

Przykład użycia:

```
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 Operacje arytmetyczne: dodawanie, odejmowanie, mnożenie, moduł z różnicą

Obrazy są macierzami, a zatem operacje arytmetyczne są dość proste – tak jak w pakiecie Matlab. Należy oczywiście pamiętać o konwersji na odpowiedni typ danych. Zwykle dobrym wyborem będzie `double`.

**Ćwiczenie 1.5** Wykonaj operacje arytmetyczne na obrazie *lena*.

1. Pobierz ze strony kursu obraz *lena*, a następnie go wczytaj za pomocą funkcji z OpenCV – dodaj ten fragment kodu do pliku, który zawiera wczytywanie obrazu *mandril*. Wykonaj konwersję do odcieni szarości. Dodaj macierze zawierające mandryla i Leny w skali szarości. Wyświetl wynik.
2. Podobnie wykonaj odjęcie i mnożenie obrazów.
3. Zaimplementuj kombinację liniową obrazów.
4. Ważną operacją jest moduł z różnicy obrazów. Można ją wykonać „ręcznie” – konwersja na odpowiedni typ, odjęcie, moduł (`abs`), konwersja na `uint8`. Alternatywa to wykorzystanie funkcji `absdiff` z OpenCV. Proszę obliczyć moduł z różnicy „szarych” wersji mandryla i Leny.

### Rozliczenie ćwiczenia

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu `.py` lub `.ipynb` w odpowiednim miejscu w zasobach kursu na UPeL. ■

**Uwaga!** Przy wyświetleniu obraz nie będzie poprawny, ponieważ jest on typu *float64 (double)*. Stosowna konwersja:

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

## 1.7 Wyliczenie histogramu

Wyliczanie histogramu można wykonać z wykorzystaniem funkcji `calcHist`. Jednak zanim do tego przejdziemy, przypomnimy sobie podstawowe struktury sterowania w Pythonie – funkcje i podprogramy. Proszę samodzielnie dokończyć poniższą funkcję wyliczającą histogram z obrazu w 256-ciu odcieniach szarości:

```
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
```

**Uwaga!** W Pythonie ważne są wcięcia, gdyż to one wyznaczają blok funkcji, czy pętli!

Histogram można wyświetlić wykorzystując funkcję `plt.hist` lub `plt.plot` z biblioteki *Matplotlib*.

Funkcja `calcHist` może policzyć histogram kilku obrazów (lub składowych), stąd jako parametry otrzymuje tablice (np. obrazów), a nie jeden obraz. Najczęściej jednak wykorzystywana jest postać:

```
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
```

Proszę sprawdzić czy histogramy uzyskane obiema metodami są takie same.

## 1.8 Wyrównywanie histogramu

Wyrównywanie histogramu to popularna i ważna operacja przetwarzania wstępnego.

1. Wyrównywanie "klasyczne" jest metodą globalną, wykonuje się na całym obrazie. W bibliotece OpenCV znajduje się gotowa funkcja do tego typu wyrównania:

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

2. Wyrównywanie CLAHE (ang. *Contrast Limited Adaptive Histogram Equalization*) – jest to metoda adaptacyjna, która poprawia warunki oświetleniowe na obrazie. Wyrównuje histogram w poszczególnych fragmentach obrazu, a nie dla całego obrazu. Metoda działa następująco:

- podział obrazu na rozłączne bloki (kwadratowe),
- wyliczanie histogramu w blokach,

- wykonanie wyrównywania histogramu, przy czym ogranicza się maksymalną „wysokość” histogramu, a nadmiar re-dystrybuuje na sąsiednie przedziały,
- interpolacja wartości pikseli na podstawie wyliczonych histogramów dla danych bloków (uwzględnia się cztery sąsiednie środki kwadratów).

Szczegóły na [Wiki](#) oraz [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)
```

**Ćwiczenie 1.6** Uruchom i porównaj obie metody wyrównywania.

#### **Rozliczenie ćwiczenia**

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu .py lub .ipynb w odpowiednim miejscu w zasobach kursu na UPeL. ■

## 1.9 Filtracja

Filtracja to bardzo ważna grupa operacji na obrazach. W ramach ćwiczenia proszę uruchomić:

- filtrację Gaussa (GaussianBlur)
- filtrację Sobela (Sobel)
- Laplasjan (Laplacian)
- medianę (medianBlur)

**Uwaga!** Pomocna będzie [dokumentacja OpenCV](#).

Proszę zwrócić uwagę również na inne dostępne funkcje:

- filtrację bilateralną,
- filtry Gabora,
- operacje morfologiczne.

**Ćwiczenie 1.7** Uruchom i porównaj wymienione metody.

#### **Rozliczenie ćwiczenia**

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu .py lub .ipynb w odpowiednim miejscu w zasobach kursu na UPeL. ■



# Laboratorium 2

<b>2</b>	<b>Detekcja pierwszoplanowych obiektów ruchomych</b>	<b>17</b>
2.1	Wczytywanie sekwencji obrazów	
2.2	Odejmowanie ramek i binaryzacja	
2.3	Operacje morfologiczne	
2.4	Indeksacja i prosta analiza	
2.5	Ewaluacja wyników detekcji obiektów pierwszoplanowych	
2.6	Przykładowe rezultaty algorytmu do detekcji ruchomych obiektów pierwszoplanowych	





## 2. Detekcja pierwszoplanowych obiektów ruchomych

### Co to są obiekty pierwszoplanowe ?

Takie, które są dla nas (dla rozważanej aplikacji) interesujące. Zwykle: ludzie, zwierzęta, samochody (lub inne pojazdy), bagaże (potencjalne bomby). Zatem definicja jest ściśle związana z docelową aplikacją.

### Czy segmentacja obiektów pierwszoplanowych to segmentacja obiektów ruchomych ?

Nie. Po pierwsze, obiekt, który się zatrzymał, nadal może być dla nas interesujący (np. stojący przed przejściem dla pieszych człowiek). Po drugie, istnieje cały szereg ruchomych elementów sceny, które nie są dla nas interesujące. Przykłady to płynąca woda, fontanna, poruszające się drzewa i krzaki itp. Należy również zauważyć, że zwykle obiekty ruchome są brane pod uwagę podczas przetwarzania obrazu. Zatem detekcję obiektów pierwszoplanowych można i warto wspomagać detekcją obiektów ruchomych.

Najprostsza metodą detekcji obiektów ruchomych to przeprowadzenie odejmowania kolejnych (sąsiednich) ramek. W ramach ćwiczenia zrealizujemy proste odejmowanie dwóch ramek, połączone z binaryzacją, indeksacją i analizą otrzymanych obiektów. Na koniec spróbujemy uznać wynik odejmowania jako rezultat segmentacji obiektów pierwszoplanowych i sprawdzimy jakość tej segmentacji.

### 2.1 Wczytywanie sekwencji obrazów

#### Ćwiczenie 2.1 Wczytywanie sekwencji obrazów

1. Ściągnij odpowiednią sekwencję testową z platformy UPeL. W ćwiczeniu skupimy się na sekwencji *pedestrians*. Wykorzystywane sekwencje pochodzą ze zbioru danych ze strony [changedetection.net](http://changedetection.net). Zbiór zawiera poetykietowane sekwencje, tzn. każda ramka

obrazu posiada maskę obiektów - maskę referencyjną (ang. *ground truth*). Na nich każdy z pikseli został przydzielony do jednej z pięciu kategorii: tło (0), cień (50), poza obszarem zainteresowania (85), nieznan ruch (170) oraz obiekty pierwszoplanowe (255) – w nawiasach podano odpowiadające poziomy szarości.

W ramach ćwiczenia interesować nas będzie tylko podział na obiekty pierwszoplanowe oraz pozostałe kategorie. Dodatkowo w folderze znajduje się maska obszaru zainteresowania (ROI) oraz plik tekstowy z przedziałem czasowym dla którego należy analizować wyniki (temporalROI.txt) – szczegóły w dalszej części ćwiczenia.

2. Wczytanie sekwencji umożliwia następujący, przykładowy kod:

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

**R** Zakłada się, że sekwencja została rozpakowana w tym samym folderze, co plik źródłowy (w innym przypadku trzeba dodać odpowiednią ścieżkę).

Należy użyć funkcji `waitKey`. Inaczej nie nastąpi odświeżenie wyświetlanego obrazka.

3. Do pętli dodaj opcję analizy co i-tej ramki – wykorzystywana funkcja *range* może mieć jako trzeci parametr: *krok*. Poeksperymentuj z jego wartością teraz (przy wyświetlaniu filmu) i później (przy detekcji obiektów ruchomych).

## 2.2 Odejmowanie ramek i binaryzacja

W celu detekcji elementu ruchomych, odejmujemy dwie kolejne ramki. Należy zaznaczyć, że dokładnie chodzi nam o obliczenie modułu z różnicy.

**Uwaga!** Aby uniknąć problemów z odejmowaniem liczb bez znaku (*uint8*), należy wykonać konwersję do typu *int* – `IG = IG.astype('int')`.

Dla uproszczenia rozważań lepiej wcześniej dokonać konwersji do odcieni szarości. Na początku trzeba „jakoś” obsłużyć pierwszą iterację. Przykładowo – wczytać pierwszą ramkę przed pętlą i uznać ją za poprzednią. Później na końcu pętli należy dodać przypisanie ramka poprzednia = ramka bieżąca. Testowo wyświetlamy wyniki odejmowania – powinny być widoczne głównie krawędzie.

Binaryzację można wykonać wykorzystując następującą składnię:

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

**Uwaga!** W takim przypadku `1*` oznacza konwersję z typu logicznego na liczbowy. Osobną kwestią jest poprawne wyświetlenie – trzeba zmienić zakres (pomnożyć przez 255) i dokonać konwersji na *uint8* (możemy być potrzebny import biblioteki *numpy*).

Próg należy dobrać, tak aby obiekty były względnie wyraźne. Proszę zwrócić uwagę na artefakty związane z kompresją.

Alternatywa to użycie funkcji wbudowanej w 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** Pierwszy argument zwracanych wartości przez funkcję `cv2.threshold` to próg binaryzacji (jest on użyteczny w przypadku stosowania automatycznego wyznaczenia progu np. metodą Otsu lub trójkątów). Szczegóły w dokumentacji OpenCV.

## 2.3 Operacje morfologiczne

**Ćwiczenie 2.2** Uzyskany obraz jest dość zaszumiony. Proszę wykonać filtrację wykorzystując erozję i dylatację (`erode` i `dilate` z OpenCV).

**Uwaga!** Celem tego etapu jest uzyskanie maksymalnie widocznej sylwetki, przy minimalnych zakłóceniach. Dla poprawy efektu warto dodać jeszcze etap filtracji medianowej (przed morfologią) oraz ew. skorygować próg binaryzacji.

## 2.4 Indeksacja i prosta analiza

W kolejnym etapie przeprowadzimy filtrację uzyskanego wyniku. Wykorzystamy indeksację (nadanie etykiet dla grup połączonych pikseli) oraz obliczanie parametrów tychże grup. Służy do tego funkcja `connectedComponentsWithStats`. Wywołanie:

```
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** Przy wyświetlaniu obrazu *labels* trzeba w odpowiedni sposób ustawić format oraz dodać skalowanie.  
Do skalowania można wykorzystać informację o liczbie znalezionych obiektów.

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

Następnie wyświetl prostokąt otaczający, pole i indeks dla największego obiektu. Poniżej znajduje się przykładowe rozwiązanie zadania. Proszę go uruchomić i ewentualnie spróbować go zoptymalizować.

```
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))

```

Komentarze do przykładowego rozwiązania:

- `stats.shape[0]` to liczba obiektów. Ponieważ funkcja zlicza też obiekt o indeksie 0 (tj. tło), w sprawdzeniu czy są obiekty warunek jest  $> 1$
- kolejne dwie linie to obliczenie indeksu maksimum z kolumny numer 4 (pole).
- uzyskany indeks należy inkrementować, ponieważ w analizie pominęliśmy element 0 (tło).
- do rysowania prostokąta otaczającego na obrazie wykorzystujemy funkcję `rectangle` z OpenCV. Składnia `"%f" % stats[pi,4]` pozwala wypisać wartość w odpowiednim formacie (f - float). Kolejne parametry to współrzędne dwóch przeciwległych wierzchołków prostokąta. Następnie kolor w formacie (B, G, R), a na końcu grubość linii. Szczegóły w dokumentacji funkcji.
- do wypisywania tekstu na obrazie służy funkcja `putText`. Do określenia pozycji pola tekstowego podaje się współrzędne lewego dolnego wierzchołka. Następnie czcionka (pełna lista w dokumentacji), rozmiar i kolor.

**Uwaga!** `I_VIS` to obraz wejściowy jeszcze przed konwersją do odcieni szarości.

**Ćwiczenie 2.3** Wykorzystaj funkcję `cv2.connectedComponentsWithStats` do indeksacji oraz oblicz parametry otrzymanych obiektów. Wyświetl prostokąt otaczający, pole oraz indeks dla największego obiektu. Na podstawie wskazówek i przykładów zamieszczonych w tekście powyżej spróbuj zoptymalizować rozwiązanie. ■

## 2.5 Ewaluacja wyników detekcji obiektów pierwszoplanowych

Aby ocenić, i to najlepiej w miarę obiektywnie, algorytm detekcji obiektów pierwszoplanowych należy wyniki przez niego zwracane tj. maskę obiektów porównać z maską referencyjną (ang. *groundtruth*). Porównywanie odbywa się na poziomie poszczególnych pikseli. Jeśli wykluczy się cienie (a tak założyliśmy na wstępie) to możliwe są cztery sytuacje:

- *TP* – wynik prawdziwie dodatni (ang. *true positive*) – piksel należący do obiektu z pierwszego planu jest wykrywany jako piksel należący do obiektu z pierwszego planu,
- *TN* – wynik prawdziwie ujemny (ang. *true negative*) – piksel należący do tła jest wykrywany jako piksel należący do tła,
- *FP* – wynik fałszywie dodatni (ang. *false positive*) – piksel należący do tła jest wykrywany jako piksel należący do obiektu z pierwszego planu,
- *FN* – wynik fałszywie ujemny (ang. *false negative*) – piksel należący do obiektu jest wykrywany jako piksel należący do tła.

Na podstawie wymienionych współczynników można policzyć szereg miar. My wykorzystamy trzy: precyzję (ang. *precision* - *P*), czułość (ang. *recall* - *R*) i tzw. *F1*. Zdefiniowane są one następująco:

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

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

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

Miara F1 jest z zakresu [0;1], przy czym im jej wartość jest większa, tym lepiej.

#### Ćwiczenie 2.4 Zaimplementuj obliczanie miar $P$ , $R$ i $F1$ .

1. W pierwszym kroku należy zdefiniować globalne liczniki  $TP$ ,  $TN$ ,  $FP$ ,  $FN$ , a po skończonej pętli obliczyć żądane wartości  $P$ ,  $R$  i  $F1$ .
2. Do pętli należy dodać wczytywanie maski referencyjnej – analogicznie jak obrazka. Jest ona dostępna w folderze *groundtruth*.
3. Kolejny krok to porównanie maski obiektów i referencyjnej. Najprostsze rozwiązanie, czyli pętla `for` po całym obrazie, w każdej iteracji sprawdzamy podobieństwo pikseli. Oczywiście nie jest to zbyt wydajne podejście. Można spróbować wykorzystać możliwości Python'a w realizacji operacji na macierzach. Utwórz odpowiednie warunki logiczne np. dla  $TP$ :

```
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
```

Przy czym obliczenia wykonujemy tylko wtedy, gdy dostępna jest poprawna mapa referencyjna. W tym celu należy sprawdzić zależność licznika ramki i wartości z pliku *temporalROI.txt* – musi się on zawierać w zakresie tam opisanym. Przykładowy kod wczytujący zakres (przy okazji proszę zwrócić uwagę na sposób obsługi plików tekstowych):

```
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. Przeprowadź obliczenia dla metody odejmowania kolejnych klatek. Zwróć uwagę na wartość  $F1$ .
5. Wykonaj obliczenia dla pozostałych dwóch sekwencji – *highway* i *office*.

#### Rozliczenie ćwiczenia

W celu rozliczenia ćwiczenia, po wykonaniu zadania zgłoś swoją gotowość prowadzącemu zajęcia. Równocześnie możesz zgłosić swoją gotowość po wykonaniu wszystkich zadań dotyczących poruszanego tematu zajęć. Po akceptacji z jego strony, umieść skrypt o rozszerzeniu `.py` lub `.ipynb` w odpowiednim miejscu w zasobach kursu na UPeL.

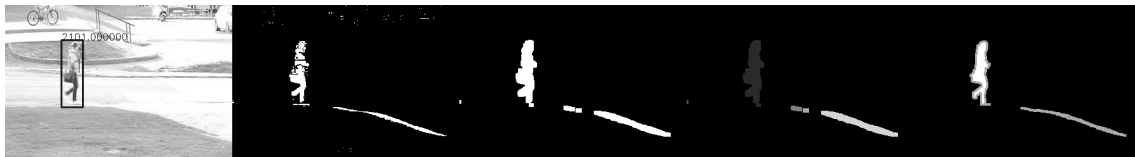
#### Informacje dotyczące kolejnych zajęć.

1. W ramach dalszych ćwiczeń będziemy poznawać kolejne algorytmy i funkcjonalności języka Python. Jednak nie kładziemy szczególnej uwagi na poznawanie języka Python, ale na wykorzystaniu w praktyce kolejnych algorytmów pojawiających się na wykładach. Przedmiot Zaawansowane Algorytmy Wizyjne nie jest kursem Pythona!
2. Przedstawione rozwiązania należy zawsze traktować jako przykładowe. Na pewno problem można rozwiązać inaczej, a czasem lepiej.

3. W kolejnych ćwiczeniach poziom szczegółowości opisu rozwiązania będzie maleć. Zachęcamy zatem do aktywnego korzystania z dokumentacji do Pythona, OpenCV i materiałów/tutoriali znajdujących się w internecie :).

## 2.6 Przykładowe rezultaty algorytmu do detekcji ruchomych obiektów pierwszoplanowych

W celu lepszej weryfikacji poszczególnych etapów zaproponowanej metody do detekcji ruchomych obiektów pierwszoplanowych przedstawiony zostanie przykładowy wynik dla obrazu o indeksie 350.



Rysunek 2.1: Przykładowy wynik poszczególnych etapów algorytmu. Od lewej obraz wejściowy z prostokątem otaczającym, obraz po binaryzacji, obraz po zastosowaniu filtracji medianowej oraz operacji morfologicznych (erode, dilate), obraz reprezentujący etykiety po indeksacji, obraz referencyjny.



# Laboratorium 3

<b>3</b>	<b>Segmentacja obiektów pierwszoplanowych</b>	<b>25</b>
3.1	Cel zajęć	
3.2	Segmentacja obiektów pierwszoplanowych	
3.3	Metody oparte o bufor próbek	
3.4	Aproksymacja średniej i mediany (tzw. sigma-delta)	
3.5	Polityka aktualizacji	
3.6	OpenCV – GMM/MOG	
3.7	OpenCV – KNN	
3.8	Przykładowe rezultaty algorytmu do segmentacji obiektów pierwszoplanowych	
3.9	Wykorzystanie sieci neuronowej do segmentacji obiektów pierwszoplanowych	
3.10	Zadania dodatkowe	





## 3. Segmentacja obiektów pierwszoplanowych

### 3.1 Cel zajęć

- zapoznanie z zagadnieniem segmentacji obiektów pierwszoplanowych oraz problemami z tym związanymi,
- implementacja prostych algorytmów modelowania tła opartych o bufor próbek – analiza ich wad i zalet,
- implementacja algorytmów średniej bieżącej oraz aproksymacji medianowej – analiza ich wad i zalet,
- poznanie metod dostępnych w *OpenCV* – Gaussian Mixture Model (zwanej też Mixture of Gaussians) oraz metody opartej o algorytm KNN (K-Nearest Neighbours, K-najbliższych sąsiadów),
- zapoznanie się z architekturą sieci neuronowej oraz przykładową aplikacją.

### 3.2 Segmentacja obiektów pierwszoplanowych

**Co to jest segmentacja obiektów pierwszoplanowych (ang. *foreground object segmentation*)**

Segmentacja to wyodrębnienie pewnej grupy obiektów obecnych na obrazie. Należy zaznaczyć, że niekoniecznie musi być ona tożsama z klasyfikacją, gdzie wyjściem jest przypisanie obiektu do konkretnej klasy: np. samochód (a nawet typ), pieszy itp. Definicja obiektu pierwszoplanowego nie jest bardzo ścisła – są to wszystkie obiekty istotne dla danej aplikacji. Przykładowo dla monitoringu wizyjnego: ludzie, samochody i może zwierzęta.

**Co to jest model tła ?**

W najprostszym przypadku jest to obraz „pustej sceny”, tj. bez interesujących nas obiektów. Uwaga. W bardziej zaawansowanych algorytmach, model tła nie ma postaci „jawnej” (nie jest to obraz) i nie można go po prostu wyświetlić (przykłady takich metod: GMM, ViBE, PBAS).

### Inicjalizacja modelu tła

W przypadku uruchomienia algorytmu (także ponownego uruchomienia), konieczne jest wykonanie inicjalizacji modelu tła tj. ustalenie wartości wszystkich parametrów (zmiennych), które stanowią ten model. W najprostszym przypadku, za początkowy model tła, przyjmuje się pierwszą ramkę z analizowanej sekwencji (pierwsze N-ramek w przypadku metod z buforem). Podejście to sprawdza się dobrze, przy założeniu, że na tej konkretnej ramce (sekwencji ramek) nie występują obiekty z pierwszego planu. W przeciwnym przypadku początkowy model będzie zawierał błędy, które będą mniej lub bardziej trudne do eliminacji w dalszym działaniu — zależy to od konkretnego algorytmu. W literaturze przedmiotu zagadnienie to określa się mianem *bootstrap*.

Bardziej zaawansowane metody inicjalizacji polegają na analizie czasowej, a nawet przestrzennej pikseli dla pewnej początkowej sekwencji. Zakłada się tutaj, że tło jest widoczne przez większość czasu oraz jest bardziej „spójne przestrzennie” niż obiekty.

### Modelowanie tła, generacja tła

Powstaje pytanie, czy nie wystarczy pobrać obrazu pustej sceny, zapisać go i używać w trakcie całego działania algorytmu? W ogólnym przypadku nie, ale zacznijmy od przypadku szczególnego. Jeśli nasz system nadzoruje pomieszczenie, w którym oświetlenie jest ściśle kontrolowane, a jakiegokolwiek uprzednio nieprzewidziane i zatwierdzone zmiany nie powinny mieć miejsca (np. zabroniona strefa wewnątrz elektrowni, skarbiec banku itp.), to najprostsze podejście ze statycznym tłem się sprawdzi.

Problemy pojawiają się w przypadku, gdy oświetlenie sceny ulega zmianie, czy to z powodu zmiany pory dnia, czy zapalenia dodatkowego światła itp. Wtedy „rzeczywiste tło” ulegnie zmianie i nasz statyczny model tła nie będzie mógł się do tej zmiany dostosować. Drugim, trudniejszym przypadkiem jest zmiana położenia elementów sceny: np. ktoś przestawi ławkę/krzesło/kwiatek. Zmiana ta nie powinna być wykrywana, a przynajmniej po dość krótkim czasie uwzględniona w modelu tła. Inaczej będzie przyczyną generowania fałszywych detekcji (znanych pod nazwą *ghost* – duch), a w konsekwencji fałszywych alarmów.

Wniosek z rozważań jest następujący. Dobry model tła powinien charakteryzować się zdolnością adaptacji do zmian na scenie. Zjawisko aktualizacji modelu określa się w literaturze mianem generacji tła lub modelowania tła.

### Pałapki przy modelowaniu tła

Rozważane zagadnienie nie jest proste. Na przestrzeni lat powstało wiele metod, ale nie jest dostępna jedna uniwersalna i skuteczna w każdych okolicznościach. W literaturze wyróżnia się następujące sytuacje „problematiczne”:

- szum na obrazie (nadal obecny pomimo znacznych postępów w technice produkcji czujników wizyjnych, szczególnie widoczny przy słabym oświetleniu), do tej kategorii warto też zaliczyć artefakty związane z kompresją cyfrowego strumienia wideo (MJPEG, H.264/265, MPEG),
- drżenie kamery – dobrym przykładem są kamery monitorujące ruch drogowy zamontowane na słupach sygnalizacji świetlnej,
- automatyczne nastawy kamery (balans bieli, korekcja ekspozycji) – np. jeśli na scenie pojawi się względnie duży jasny obiekt, to „automatyka” spowoduje zmianę ekspozycji i/lub balansu bieli dla całej sceny, co zwykle kończy się fatalnie dla algorytmu detekcji obiektów,
- pora dnia (ogólnie płynne zmiany oświetlenia),
- nagłe zmiany oświetlenia (włączenie światła w pomieszczeniu, słoneczno-pochmurny dzień),
- obecność obiektów podczas inicjalizacji modelu tła (skomentowane wcześniej),

- ruchome elementy tła (skomentowane wcześniej),
- kamuflaż — duże podobieństwo obiektów do tła pod względem koloru lub tekstury (tu najnowszy trend to wykorzystanie informacji 3D),
- poruszone obiekty w tle (krzesła itp.)

Ostatnie i chyba najtrudniejsze zagadnienia to wtapianie się obiektów z pierwszego planu w tło oraz tak zwane duchy (ang. *ghost*). Oba zjawiska są ze sobą połączone. Rozważmy następującą sytuację. Samochód wjeżdża na parking i zatrzymuje się. Początkowo jest wykrywany jako obiekt, po ale po pewnym czasie (zależnym od metody) stanie się elementem tła. Następnie samochód odjedzie. Puste miejsce po nim będzie wykrywane jako obiekt, gdyż znacząco różni się od bieżącej ramki. Powstaje *ghost*, czyli obiekt, który nie istnieje na bieżącej ramce obrazu, a tylko w modelu tła.

### Polityka aktualizacji

Jak zostało wcześniej stwierdzone, podstawą modelowania (generacji) tła, jest aktualizacja modelu tła. Wyróżnia się dwa skrajne podejścia do zagadnienia: konserwatywne i liberalne. W pierwszym przypadku aktualizujemy tylko te piksele, które zostały sklasyfikowane jako tło. W drugim aktualizacja wykonywana jest dla wszystkich pikseli. Oczywiście występuje też cały szereg podejść pośrednich.

Warto zwrócić uwagę na skutki obu polityk. W przypadku konserwatywnej łatwo narażamy się na sytuację, w której błędna klasyfikacja będzie podtrzymywana w nieskończoność, gdyż ten obszar nigdy nie będzie aktualizowany. Z drugiej strony podejście liberalne spowoduje wtapianie obiektów z pierwszego planu do modelu tła i może prowadzić do powstawania *ghost'ów* lub smug ciągnących się za obiektami.

### Cienie

Cienie są bardzo specyficzną kategorią w zagadnieniach detekcji obiektów. Człowiek podświadomie cienie pomija, tj. jak analizujemy scenę to raczej specjalnie nie zwracamy uwagi na to czy np. sylwetka rzuca cień, czy nie. W szczególności dotyczy to cieni słabo widocznych.

W systemie detekcji cień spełnia wszystkie założenia obiektu z pierwszego planu i raczej ciężko go uznać za element tła. Rozważania dotyczą cieni rzucanych przez obiekty pierwszoplanowe. Problem polega na tym, że z punktu widzenia działania dalszych etapów tj. np. klasyfikacji obiektów, stanowi on zakłócenie – mocno utrudnia analizę np. poprzez zmianę kształtu obiektu lub łączenie obiektów.

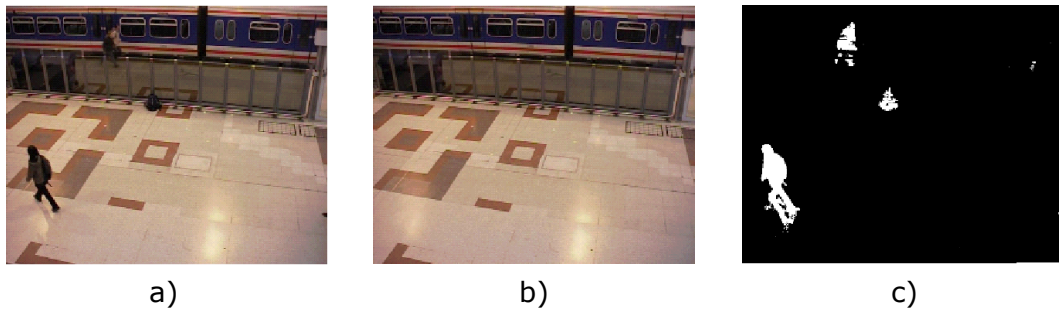
Z wyżej wymienionych powodów prowadzone są badania nad detekcją i eliminacją cieni. Na potrzeby niniejszego ćwiczenia należy zapamiętać, że cień powinien stanowić osobną kategorię - nie jest on ani obiektem pierwszoplanowym, ani elementem tła.

### Przykład

Na rysunku ?? przedstawiono przykład segmentacji obiektów z wykorzystaniem modelowania tła. Warto zwrócić na wykryty cień pod nogami osoby po lewej stronie ekranu oraz osobę za barierką (raczej mało widoczną na bieżącej ramce).

## 3.3 Metody oparte o bufor próbek

W ramach ćwiczenia zaprezentowane zostaną dwie metody: średnia z bufora oraz mediana z bufora. Rozmiar bufora przyjmiemy na  $N = 60$  próbek. W każdej iteracji algorytmu należy usunąć ostatnią ramkę z bufora, dodać bieżącą (bufor działa na zasadzie kolejki FIFO) oraz obliczyć średnią lub medianę z bufora.



Rysunek 3.1: Przykład segmentacji obiektów pierwszoplanowych z wykorzystaniem modelowania tła. a) bieżąca ramka, b) model tła, c) maska obiektów pierwszoplanowych. Źródło: sekwencja PETS 2006

1. Na początku należy zadeklarować bufor o rozmiarze  $N \times YY \times XX$  (polecenie np. `zeros`), gdzie  $YY$  i  $XX$  to odpowiednio wysokość i szerokość ramki z sekwencji *pedestrians*.

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

Należy też zainicjalizować licznik dla bufora (niech się nazywa np. `iN`) wartością 0. Parametr `iN` jest jak wskaźnik, który wskazuje na miejsce w buforze, z którego należy usunąć ostatnią ramkę oraz przypisać obecną ramkę.

2. Obsługa bufora powinna być następująca. W pętli pod adresem `iN` należy zapisać bieżącą ramkę w skali szarości.

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

Następnie licznik `iN` należy inkrementować i sprawdzać czy nie osiąga rozmiaru bufora ( $N$ ). Jeśli tak, to trzeba go ustawić na 0.

3. Obliczenie średniej lub mediany realizuje się funkcjami `mean` lub `median` z biblioteki *numpy*. Jako parametr podaje się wymiar, dla którego ma być liczona wartość (pamiętać o indeksowaniu od zera). Aby wszystko działało poprawnie, należy dokonać konwersji wyniku na `uint8`.
4. Ostatecznie realizujemy odejmowanie tła tj. od bieżącej sceny odejmujemy model, a wynik binaryzujemy – analogicznie jak przy różnicy sąsiednich ramek. Również wykorzystujemy filtrację medianową maski obiektów lub/i operacje morfologiczne.
5. Wykorzystując stworzony w ramach poprzedniego ćwiczenia kod, porównaj działanie metody ze średnią i medianą. Zanotuj wartości wskaźnika  $F1$  dla obu przypadków. Mediana może liczyć się przez dłuższą chwilę. Zastanów się dlaczego mediana działa lepiej. Sprawdź działanie na innych sekwencjach – w szczególności dla sekwencji *office*. Zaobserwuj zjawisko smużenia oraz wtapiania się sylwetki do tła oraz powstania *ghosta*.

**Ćwiczenie 3.1** Zaimplementuj algorytm na podstawie powyższego opisu. ■

### 3.4 Aproksymacja średniej i mediany (tzw. sigma-delta)

Użycie bufora na próbki jest wymaga sporych zasobów pamięciowych. Poza tym, o ile średnią da się policzyć w prosty sposób (proszę się zastanowić jak można zaktualizować średnią z bufora bez ponownego liczenia dla wszystkich elementów), to już z medianą są problemy i bardzo szybkie algorytmy jej wyznaczania nie istnieją (aczkolwiek nie trzeba sortować wszystkich elementów w każdej

iteracji). Dlatego chętniej wykorzystuje się metody, które nie wymagają bufora. W przypadku aproksymacji średniej wykorzystuje się zależność:

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

gdzie:  $I_N$  – bieżąca ramka,  $BG_N$  – model tła,  $\alpha$  – parametr wagowy, zwykle 0.01 – 0.05, choć wartość zależy od konkretnej aplikacji.

Aproksymację mediany uzyskuje się wykorzystując zależność:

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

### Ćwiczenie 3.2 Zaimplementuj obie metody.

1. Jako pierwszy model tła przyjmij pierwszą ramkę z sekwencji. Zanotuj wartość wskaźnika F1 dla tych dwóch metod (parametr  $\alpha$  ustal na 0.01). Zaobserwuj czas działania.
2. Poeksperymentuj z wartością parametru  $\alpha$ . Zobacz jaki ona ma wpływ na model tła.

Dla metody średniej bieżącej bardzo ważne jest, aby model tła był zmiennoprzecinkowy (*float64*).

W przypadku wzoru 3.1 uniknięcie stosowania pętli po całym obrazie jest oczywiste. Dla zależności 3.2 również jest to możliwe, ale wymaga chwili zastanowienia. Warto korzystać z dokumentacji *numpy*. Dodatkowo proszę zauważyć, że w Pythonie można wykonywać działania arytmetyczne na wartościach boolowskich – `False == 0` natomiast `True == 1`.

## 3.5 Polityka aktualizacji

Jak dotąd stosowaliśmy liberalną politykę aktualizacji – aktualizowaliśmy wszystko. Spróbujemy teraz wykorzystać podejście konserwatywne.

### Ćwiczenie 3.3 Polityka aktualizacji.

1. Dla wybranej metody zaimplementuj konserwatywne podejście do aktualizacji. Sprawdź jego działanie.
 

**R** Wystarczy zapamiętać poprzednią maskę obiektów i odpowiednio wykorzystać ją w procedurze aktualizacji.
2. Zwróć uwagę na wartość wskaźnika F1 oraz na model tła. Czy pojawiły się w nim jakieś błędy ?

## 3.6 OpenCV – GMM/MOG

W bibliotece OpenCV dostępna jest jedna z najpopularniejszych metod segmentacji obiektów pierwszoplanowych: Gaussian Mixture Models (GMM) lub Mixture of Gaussians (MoG) — obie nazwy występują równolegle w literaturze. W największym skrócie: scena jest modelowana za pomocą kilku rozkładów Gaussa (średnia jasności (koloru) i odchylenie standardowe). Każdy rozkład opisany jest również przez wagę, która oddaje to jak często był on używany (obserwowany na scenie – prawdopodobieństwo wystąpienia). Rozkłady o największych wagach stanowią tło.

Segmentacja polega na obliczaniu odległości pomiędzy bieżącym pikselem, a każdym z rozkładów. Następnie jeśli piksel jest „podobny” do rozkładu uznanego za tło, to jest klasyfikowany jako tło, w innym przypadku uznany zostaje za obiekt. Model tła podlega również aktualizacji w sposób zbliżony do równania 3.1. Zainteresowane osoby odsyłamy do literatury np. [do artykułu](#).

**Ćwiczenie 3.4** Metoda ta jest przykładem algorytmu wielowariantowego – model tła ma kilka możliwych reprezentacji (wariantów).

1. Otwórz dokumentację do [OpenCV](#). Przejdź do *Video Analysis->Motion Analysis*. Interesuje nas klasa `BackgroundSubtractorMOG2`. Aby utworzyć obiekt tej klasy w Python3 należy użyć `createBackgroundSubtractorMOG2`. Korzystając z tego obiektu w pętli dla każdego obrazu wykorzystujemy jego metodę `apply`.
2. Na początek uruchomimy kod z wartościami domyślnymi.
3. Proszę poeksperymentować z parametrami: *history*, *varThreshold* oraz wyłączyć detekcję cieni. W metodzie `apply()` można też „ręcznie” ustalić współczynnik uczenia – *learningRate*. W pakiecie OpenCV dostępna jest wersja GMM nieco inna niż oryginalna – między innymi wyposażona w moduł detekcji cieni oraz dynamicznie zmienianą liczbę rozkładów Gaussa. Odnośniki do stosownych artykułów dostępne są w dokumentacji OpenCV.
4. Wyznacz parametr F1 (dla metody bez detekcji cieni). Zaobserwuj jak działa metoda. Zwróć uwagę, że nie da się „wyświetlić” modelu tła.

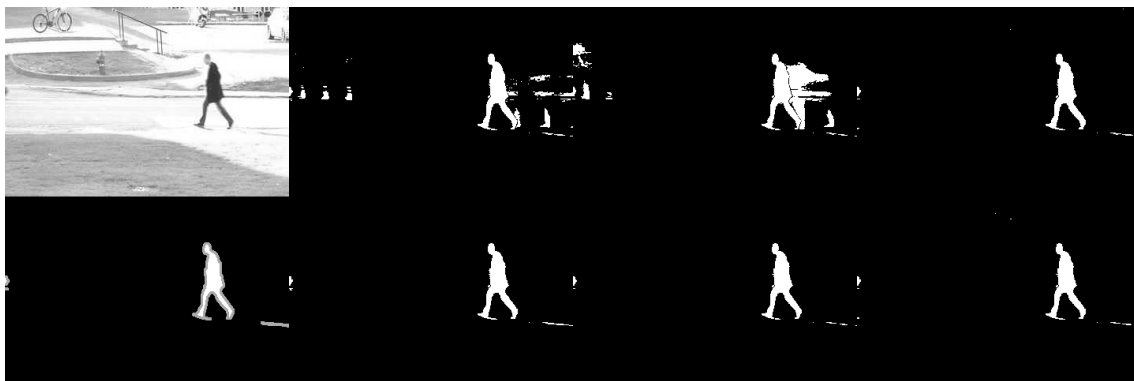
### 3.7 OpenCV – KNN

Druga z metod dostępnych w OpenCV to KNN.

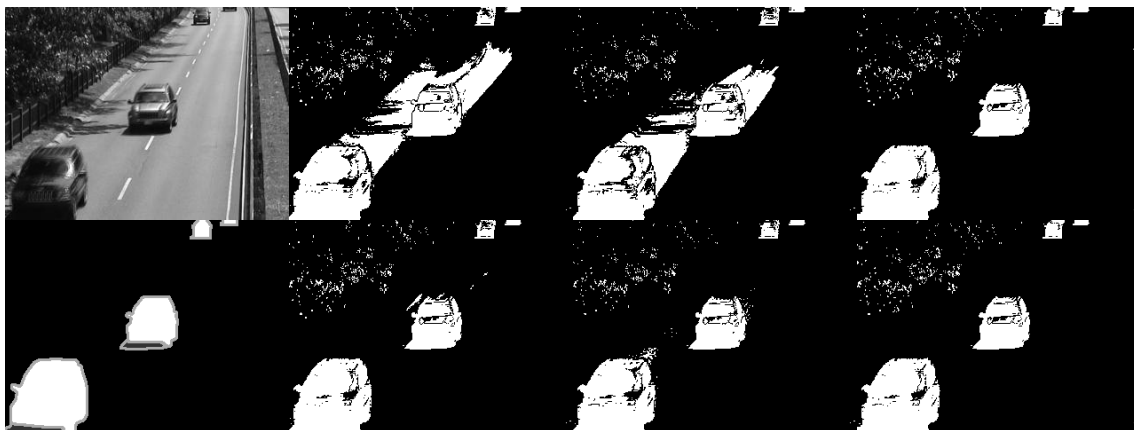
**Ćwiczenie 3.5** Proszę uruchomić metodę KNN i dokonać jej ewaluacji.

### 3.8 Przykładowe rezultaty algorytmu do segmentacji obiektów pierwszoplanowych

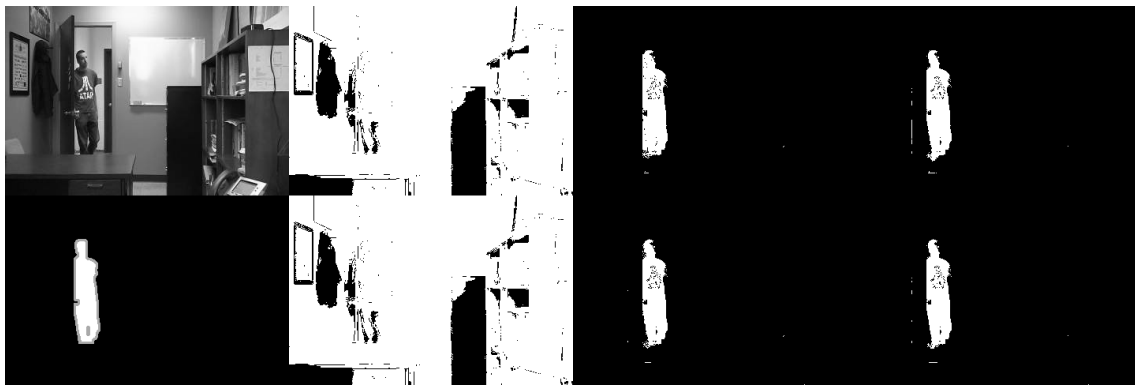
W celu lepszej weryfikacji poszczególnych etapów zaproponowanych metod do segmentacji obiektów pierwszoplanowych przedstawione zostaną przykładowe wyniki dla wybranych ramek obrazu.



Rysunek 3.2: Przykładowy wynik poszczególnych wersji algorytmu do segmentacji obiektów pierwszoplanowych dla zbioru *pedestrian*. Indeks ramki to 600. Pierwsza od lewej kolumna to obraz wejściowy wraz z obrazem referencyjnym. Pierwszy wiersz dotyczy algorytmów wykorzystujących średnią, drugi wiersz wykorzystuje medianę. Kolejno od drugiej kolumny prezentowany jest wynik algorytmu: liberalna polityka aktualizacji, następnie aproksymacja funkcji z liberalną polityką aktualizacji oraz aproksymacja funkcji z konserwatywną polityką aktualizacji.



Rysunek 3.3: Przykładowy wynik poszczególnych wersji algorytmu do segmentacji obiektów pierwszoplanowych dla zbioru *highway*. Indeks ramki to 1200. Pierwsza od lewej kolumna to obraz wejściowy wraz z obrazem referencyjnym. Pierwszy wiersz dotyczy algorytmów wykorzystujących średnią, drugi wiersz wykorzystuje medianę. Kolejno od drugiej kolumny prezentowany jest wynik algorytmu: liberalna polityka aktualizacji, następnie aproksymacja funkcji z liberalną polityką aktualizacji oraz aproksymacja funkcji z konserwatywną polityką aktualizacji.



Rysunek 3.4: Przykładowy wynik poszczególnych wersji algorytmu do segmentacji obiektów pierwszoplanowych dla zbioru *office*. Indeks ramki to 600. Pierwsza od lewej kolumna to obraz wejściowy wraz z obrazem referencyjnym. Pierwszy wiersz dotyczy algorytmów wykorzystujących średnią, drugi wiersz wykorzystuje medianę. Kolejno od drugiej kolumny prezentowany jest wynik algorytmu: liberalna polityka aktualizacji, następnie aproksymacja funkcji z liberalną polityką aktualizacji oraz aproksymacja funkcji z konserwatywną polityką aktualizacji.



### 3.9 Wykorzystanie sieci neuronowej do segmentacji obiektów pierwszoplanowych

Do segmentacji obiektów pierwszoplanowych można wykorzystać również metody oparte na architekturze sieci neuronowej. Przykładem takiego rozwiązania jest praca<sup>1</sup>. Zaproponowanym podejściem jest wykorzystanie konwolucyjnej sieci neuronowej zwanej BSUV-Net (ang. Background Subtraction Unseen Videos Net). Architektura bazuje na strukturze U-Net z połączeniami resztkowymi (ang. residual connections), która dzieli się na enkoder oraz dekoder. Wejście to konkatenacja kilku obrazów, tj. bieżąca ramka i dwie ramki tła w różnych odstępach czasowych wraz z ich semantycznymi mapami segmentacji. Wyjściem jest maska zawierająca tylko obiekty pierwszoplanowe.

Szczegóły odnośnie tej sieci neuronowej znajdziecie w artykule: <https://arxiv.org/abs/1907.11371>. Architektura z przykładowymi danymi oraz wytrenowanymi modelami sieci jest udostępniona na GitHubie: <https://github.com/ozantezcan/BSUV-Net-inference>.

**Ćwiczenie 3.6** Uruchom konwolucyjną sieć neuronową BSUV-Net.

- Ściągnij z platformy UPEL wszystkie pliki do sieci neuronowej.
- Wykorzystana zostanie baza danych pedestrians, ale odpowiednio przeskalowana do rozmiaru wejścia sieci oraz przekonwertowana na sekwencje wideo,
- Otwórz plik `infer_config_manualBG.py` oraz zmodyfikuj ścieżki do:
  - w klasie `SemanticSegmentation` podaj ścieżkę absolutną do folderu `segmentation`
  - zmienna `root_path`,
  - w klasie `BSUVNet` podaj ścieżkę do modelu sieci `BSUV-Net-2.0.mdl` w folderze `trained_models` – zmienna `model_path`,
  - w klasie `BSUVNet` podaj ścieżkę do obrazu, który zawiera jedynie tło – pierwsze zdjęcie ze zbioru `pedestrians` – zmienna `empty_bg_path`
- W pliku `inference.py` podaj ścieżkę do wejścia sieci, czyli sekwencja wideo `pedestrians` – zmienna `inp_path` oraz ścieżkę do miejsca gdzie ma być zapisane wyjście sieci wraz z nazwą pliku – zmienna `out_path`.

### 3.10 Zadania dodatkowe

#### 3.10.1 Inicjalizacja modelu tła w warunkach obecności obiektów pierwszoplanowych na scenie

**Ćwiczenie 3.7 Zadanie dodatkowe 1**

Jeśli weźmiemy metodę, w której zastosowaliśmy konserwatywne podejście do aktualizacji oraz rozpoczniemy analizę sekwencji nie od ramki nr 1, a od 1000 to zobaczymy w praktyce wady założenia, że pierwsza ramka z sekwencji jest początkowym modelem tła (*ghosty* itp.) Zaproponuj rozwiązanie problemu. W aplikacji należy zademonstrować przewagę własnego podejścia nad inicjalizacją na podstawie pierwszej ramki.

Na rysunku 3.5 przedstawiono przykładowe rozwiązanie problemu.

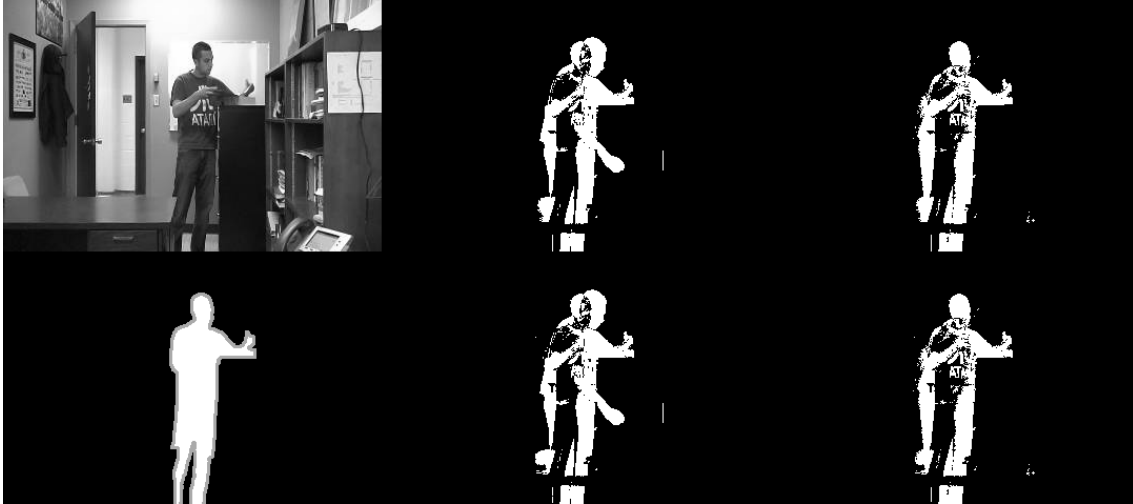


Dobrym pomysłem wydaje się zbuforowanie np. 30 lub więcej ramek oraz przeprowadzenie analizy częstotliwości występowania każdego odcienia jasności w sekwencji – niezależnie dla

<sup>1</sup>Tezcan, 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>

każdego piksela. Za tło należy uznać, ten co był obserwowany najczęściej. Inne, skuteczne pomysły, są również mile widziane.

W zadaniu można wykorzystać inną sekwencję.



Rysunek 3.5: Przykładowy wynik proponowanego rozwiązania dla zbioru *office*. Indeks ramki to 800. Pierwsza od lewej kolumna to obraz wejściowy wraz z obrazem referencyjnym. Pierwszy wiersz dotyczy algorytmów wykorzystujących średnią, drugi wiersz wykorzystuje medianę. Kolejno od drugiej kolumny prezentowany jest wynik algorytmu: aproksymacja funkcji z konserwatywną polityką aktualizacji bez buforu inicjalizacyjnego, aproksymacja funkcji z konserwatywną polityką aktualizacji z buforem inicjalizacyjnym.

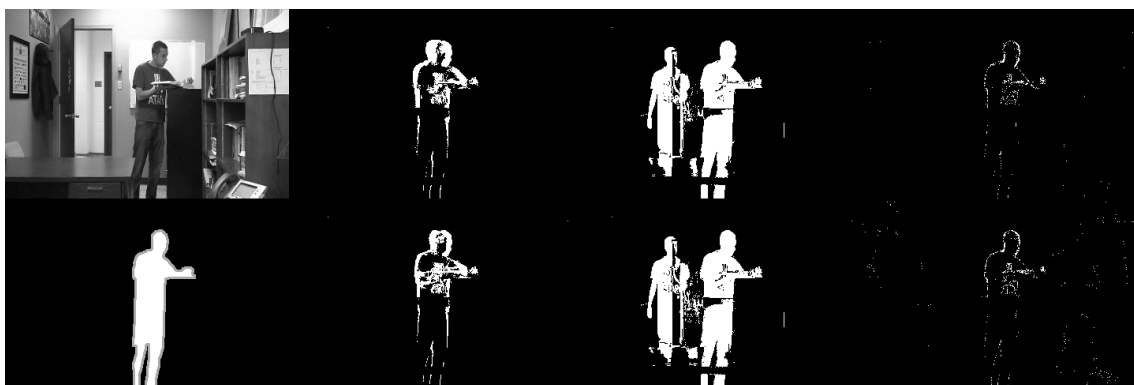
### 3.10.2 Implementacja metod ViBE i PBAS

#### Ćwiczenie 3.8 Zadanie dodatkowe 2

Na stronie kursu dostępne są artykuły, w których opisano metody ViBE i PBAS. Zadanie polega na ich implementacji oraz ewaluacji. Proszę zachować kolejność, gdyż PBAS stanowi rozszerzenie ViBE.

Na rysunku 3.6 przedstawiono przykładowe rozwiązanie problemu.





Rysunek 3.6: Przykładowy wynik proponowanego rozwiązania dla zbioru *office*. Indeks ramki to 1000. Pierwsza od lewej kolumna to obraz wejściowy wraz z obrazem referencyjnym. Następnie aproksymacja średniej oraz mediany z liberalną polityką aktualizacji, aproksymacja średniej oraz mediany z konserwatywną polityką aktualizacji oraz ostatnia kolumna to algorytm ViBe (na górze) i algorytm PBAS (na dole).