UNIVERSITY OF PADOVA

Computer Vision

Intermediate project report

Marek Tatýrek
Mateusz Miroslaw Lis

Academic Year 2024-2025

Contents

1 Introduction

1.1 Notes

To make code more clear, we use:

- `snake_case` for variables
- `PascalCase` for classes
- `camelCase` for functions

2 Overview

We decided to use class design, where classes are logicaly separated depending on task. Dataflow between classes is provided by several user structures. Data are processed squentialy. Unfortunately used dataflow is not that clear because we havent been able to fully follow originally defined structure. We also used Doxygen documentation so for detailed description you can check it here.

For loading data we have two classes `ImageLoader()` and `VideoLoader()`, derived from class `InputSource()`. From user side the classes have common interface and usage is the same. Important is method `hasNextFrame()`, used for driving the while loop in the main, returning true in case there is frame or image that we can read. Class is returning timestamp, cv::Mat with current sample.

As was already said, tasks are performed sequentialy, which is convenient, because we are reading many frames and to store them in the buffer, it will be very memory demanding. Also from the same reason we have to take care about not making deep copies.

We have `Preprocessing()` class, for editing the image before actual detection of haar features.

For detecting features in the images we are using class `HaarDetector()` using haar cascades for finding desired patterns in the images. In our case faces and eyes.

From the detected features we need to make evaluation and decide which shot type are we having. For this purpouse we have class `FeatureEvaluator()` that outputs structure `classification_result` with information if is curent sample wide shot, medium or close up.

Because we want to make statictic from the data, we made `FilmStatistics()` class. It makes sense to use this class only on video data. At the init we provide configuration structure `FilmStatisticsEvalConfig`, with many settings. We can export time sequences to `.csv` file or we can use getters to get the time sequences and use them in the code.

For graphical output of the statstic data we have `ResultDisplayer()` class, which is inputting all data types got from `FilmStatistics()` getters and returning `cv::mat` with plots.

3 Implementation details

3.1 FileLoader

In this file we have four classes.

Class `InputSource` is parent class for `ImageLoader` and `VideoLoader` It serves to provide easy interface to work with image and video files. In both cases we aiming to process data in series, that means frame by frame. Both of these classes have same user interface made of:

- `hasNextFrame()` - returns if there is one more frame to process, used for driving while loop.
- `nextFrame()` - returns next frame in the sequence.
- `getCurrentTimestamp`() - returns current timestamp in ms.

For preprocessing we have class `Preprocessing()`. We only implemented processing methods that we found beneficial. The class contains this methods:

- `LoadFrame()`
- `resizeImage()`
- `toGrayscale()`
- `equalizeHistogram()`
- `GetProcessedImage()`

Names on methods are describing function sufficiently.

3.2 SceneChangeDetect

This file contains single function for optimizing processing of Viola & Johnes detector. As an input we take two frames, we compute how much they are different and return boolean value. Because Viola & Johnes is computationaly quite expansive, we want to proccess the frame by it, only when it is really needed.

We have two parameters:

- `pixelDiffThreshold` - Difference between two pixel at same position in input frames.
- `threshold` - Percentage of pixels, that reach `pixelDiffThreshold` parameter value.

Algorithm computes difference between all pixel at apropriate positions and make boolean array. Then it computes percetage of pixels that reaches the `pixelDiffThreshold` and compare the value with `threshold`. If the computed value reaches the treshold it returns true.

3.3 TestDatasetEval

This simple class is designed for computing accuracy of the algorithm on the test dataset. That means at image dataset. Since we have test datased sorted in the folders, task is pretty simple. We have class `TestDatasetEval`. with constructor we pass `ShotType`, which we want to test with this class. Then we have two methods:

- `check()` Is used in while to add currently classified image ShotType to evaluation.
- `getEvalResults()` After all images are processed we call this method and it returns accuracy.

3.4 FilmStatistics

Class is mainly designed for video. As input it takes class probabilities and it outputs `.csv` file with statictical data. Constructor takes one argument - structure `FilmStatisticsEvalConfig`, which is used to set up all the parameters for class.

We have 4 public methods:

- `addConfigurationStruct()` - Serves to change configuration setings.
- `addFrameResult()` - Add frame to final evaluation.
- `exportToCSV()` - Exports computed statistical timelines to `csv` file for future processing.
- `printSummary()` - Prints basic statistical summary.

It is important to mention what specificaly class does. It takes input probability. We can set oversapling or skipping to input data to filter random noise etc. Class computes entropy and entropy variance, it detects cuts in the scenes and clasify most probable `ShotType`. We can set different window sizes for oversampling window, entropy window and entropy variance window. We can also reduce output data flow if we dont want to have informations from each single shot and we reduce size of output `csv` file. All of these parameters are settable in `FilmStatisticsEvalConfig`.

Since we are using sliding windows for signal processing, we introduce delay to the system, which grows as we setting bigger window sizes. Delay is different for each statistical parameter timeseries. Class handle this situation and starts outputing data when all of the parameters are valid, with correct timestamp.

3.5 Main video

Puropuse of this main is to classify video. At first we initialize all the classes and time measurements. Then we run main while loop, until we run out of frames. We can see the main loop on scheme 1

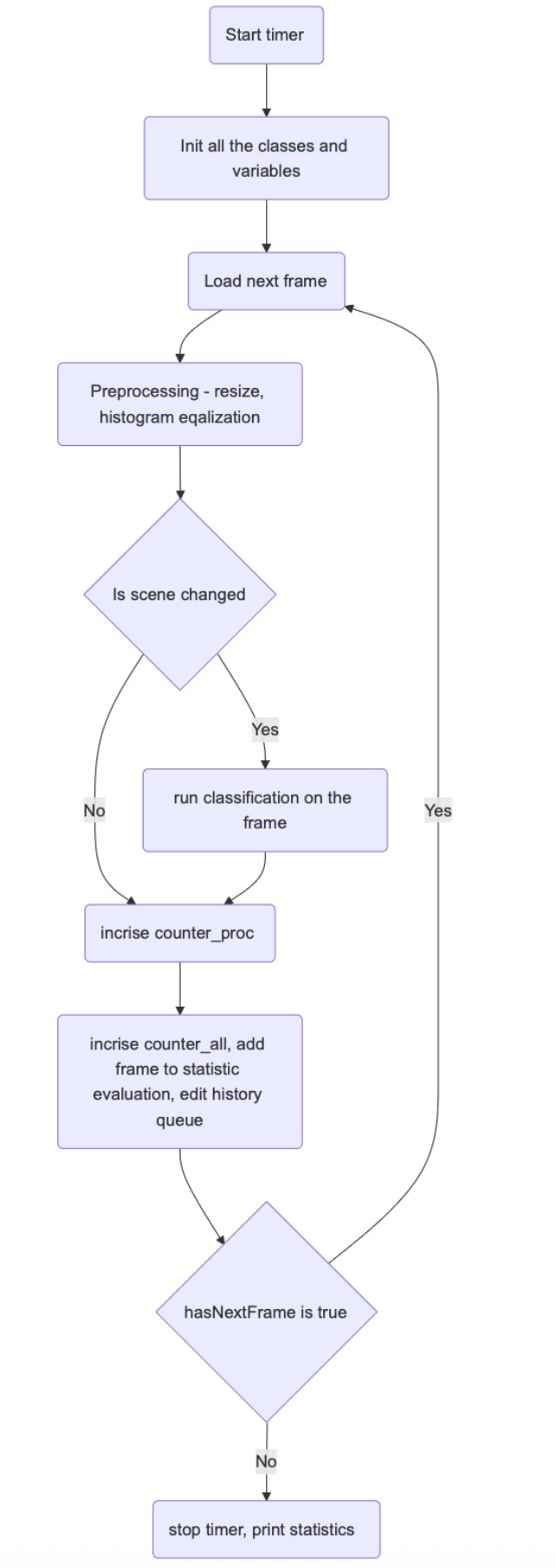


Figure 1: Classification with Viola&Johnes

4 Evaluation

4.1 Quantitative Evaluation

Evaluating our classifier system on the test dataset can be done by running `main_eval.cpp` from the submitted archive. Evaluation is done by computing one-vs-all precision, recall, F1 score, accuracy and the confusion matrix.

Shot Type	Precision	Recall	F1 Score	Accuracy
Close-up	0.8889	0.8000	0.8421	0.9000
Medium	0.8000	0.8000	0.8000	0.8667
Wide	0.7273	0.8000	0.7619	0.8333

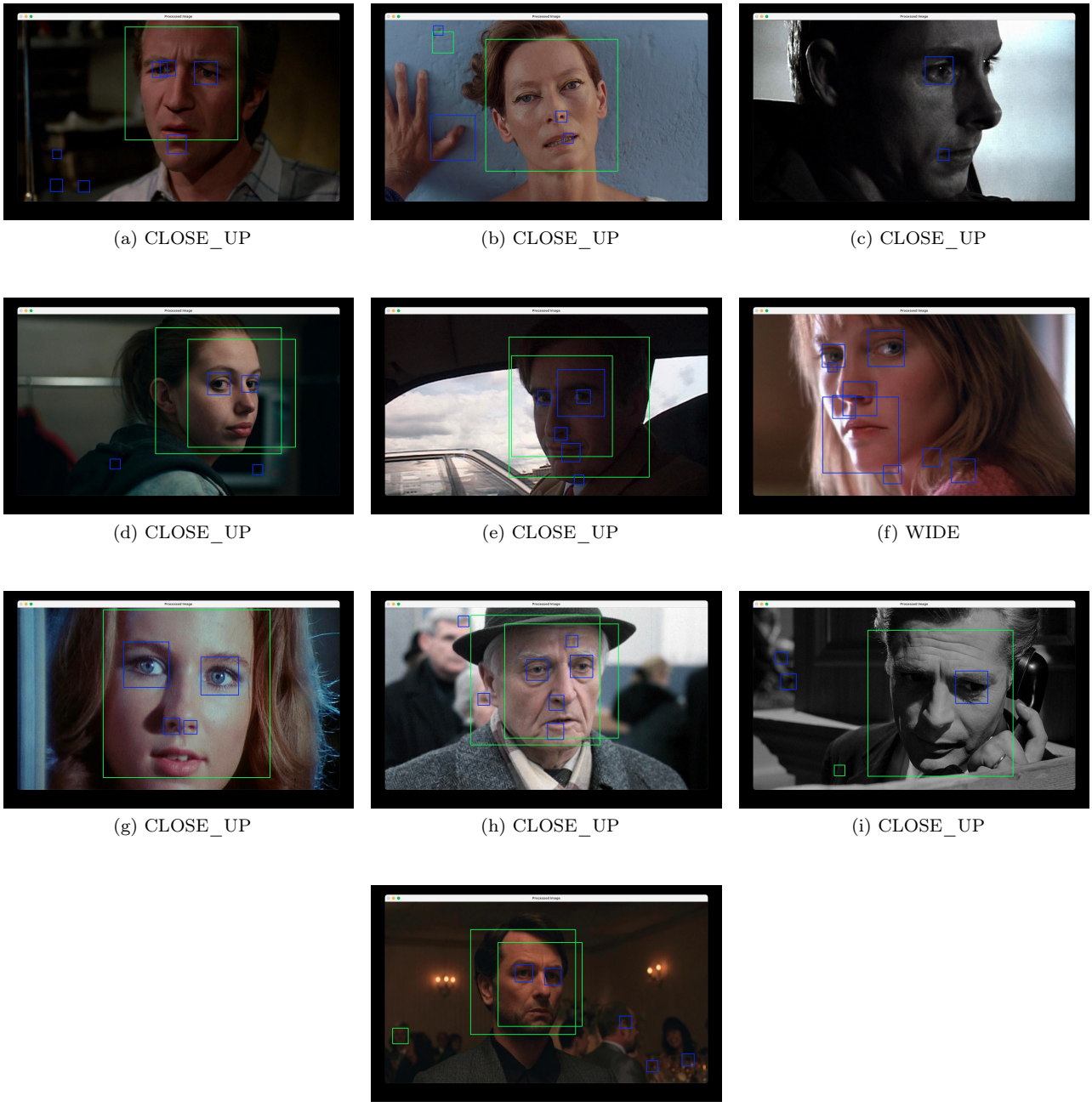
Table 1: Per-class performance metrics for shot type classification

Overall Accuracy: 0.8000

Actual \ Predicted	Close-up	Medium	Wide
Close-up	8	0	2
Medium	1	8	1
Wide	0	2	8

Table 2: Confusion matrix of the shot type classifier

4.1.1 Close up shots



(a) CLOSE_UP (b) CLOSE_UP (c) CLOSE_UP

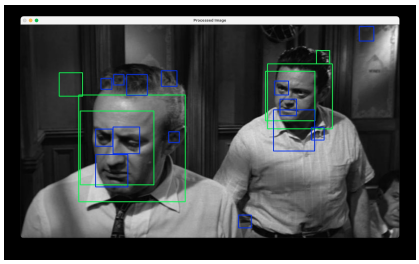
(d) CLOSE_UP (e) CLOSE_UP (f) WIDE

(g) CLOSE_UP (h) CLOSE_UP (i) CLOSE_UP

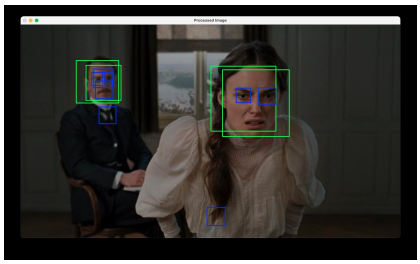
(j) CLOSE_UP

Figure 2: Examples of close-up shots with classification

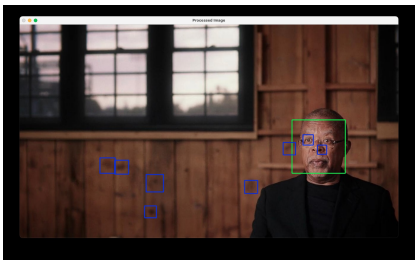
4.1.2 Medium shots



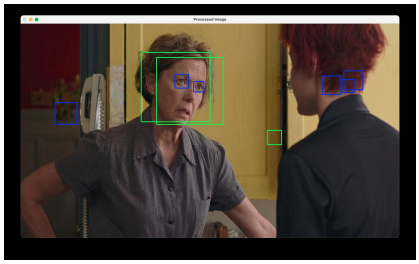
(a) MEDIUM



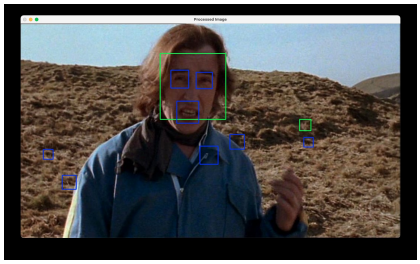
(b) MEDIUM



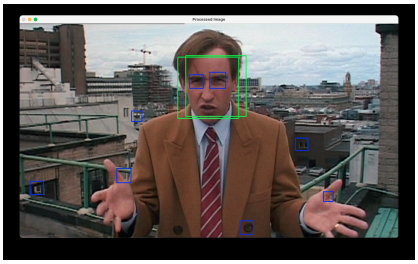
(c) MEDIUM



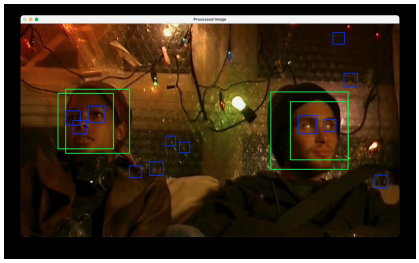
(d) CLOSE_UP



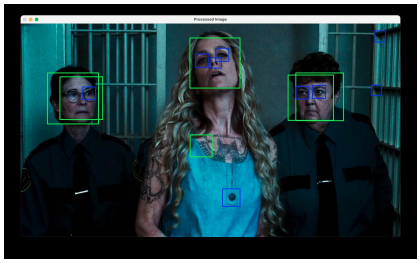
(e) MEDIUM



(f) MEDIUM

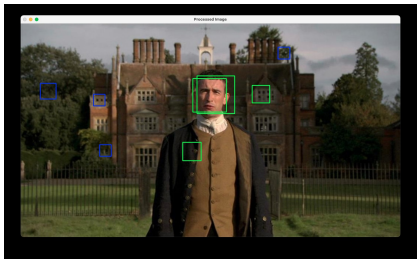


(g) MEDIUM



(h) MEDIUM

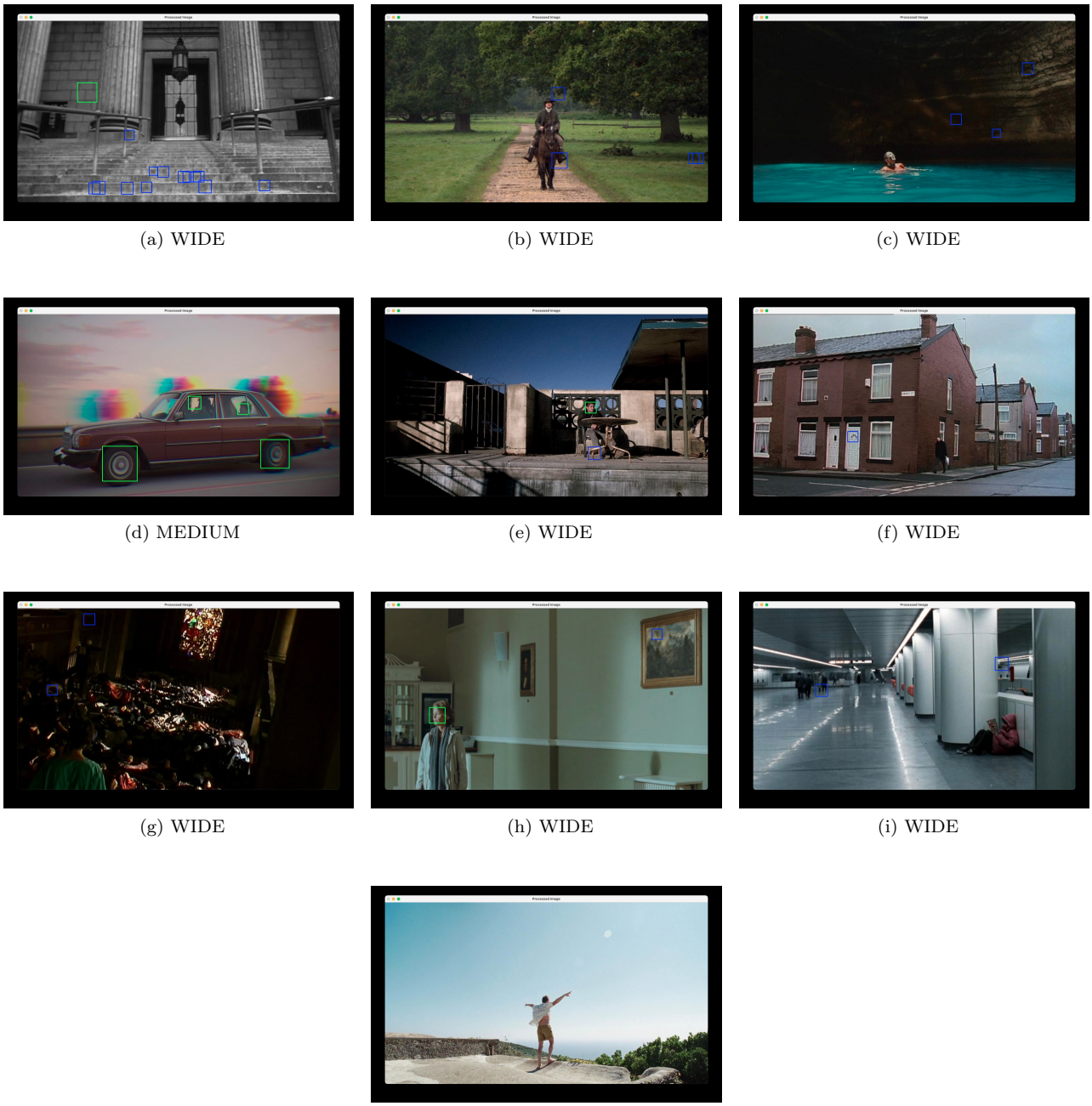
(i) WIDE



(j) MEDIUM

Figure 3: Examples of medium shots with classification

4.1.3 Wide shots



(a) WIDE (b) WIDE (c) WIDE

(d) MEDIUM (e) WIDE (f) WIDE

(g) WIDE (h) WIDE (i) WIDE

(j) WIDE

Figure 4: Examples of wide shots with classification

4.2 Video Evaluation

With video we have more parameters to evaluate such as detection times and we can also have a look on some statistics.

As testing video we used triler for Avengers [1]. We choose this type of video, because we wanted to get diverse dataset and triler are full of cuts and differents shot types. We also aimed fort short video because altrough Viola&Johnes is effective algorithm, appliing it on the video was computationally expansoive and takes significan amount of time. The sample video duration is 2:24.

We used time measurements for optimizing our algorithm, but as every machine is unique, we cant provide processing time as metric. So processing time is only to imagine how long it to compute. For reference tests were ran on Macbook 14 pro 2021 (8 CPU, 14 GPU).

But we can demonstrate here how well is our optimizer working

4.2.1 Accuracy

We are gonna analyze first 30 seconds of our video. As reference we have manually classified data. It is important to mention the setup of algorithm, we can see it at pricture 5.

Parameters of optimizer are described at 3.2, resolution parameters are for resizing the image and last 3 treshold are tresholds for classification. We also use histogram equalization in preprocessing stage.

```
constexpr int OPTIMIZER_DISTANCE_BETWEEN_SAMPLES = 3;

constexpr int RESOLUTION_WIDTH = 1280;
constexpr int RESOLUTION_HEIGHT = 720;

constexpr int PIXEL_DIFF_TRESHOLD = 60;
constexpr double IMAGE_CHANGE_RATIO = 0.5;

int constexpr WIDE_THRESHOLD = 3000;
int constexpr CLOSEUP_THRESHOLD = 50000;
int constexpr EYE_THRESHOLD = 2000;
```

Figure 5: Algorithm setup for video classification

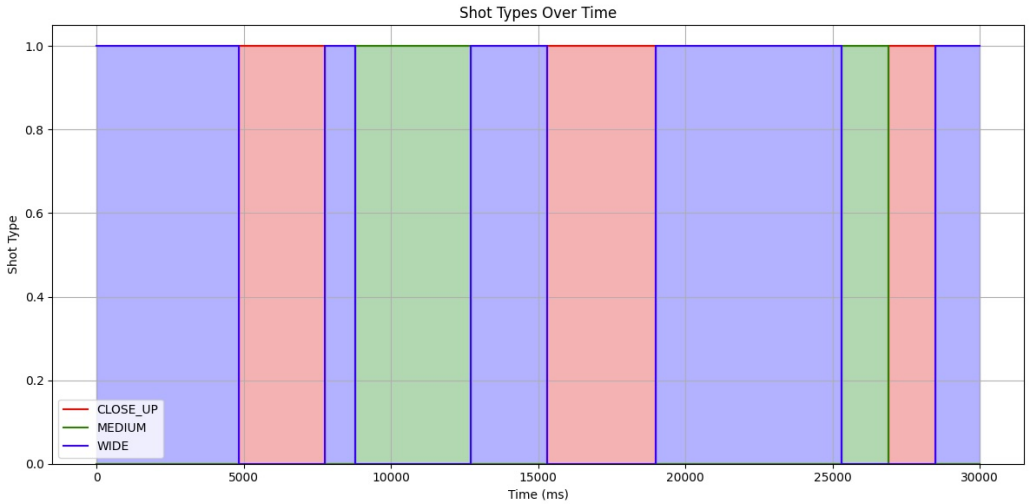


Figure 6: Manually classified video [1]

Figure 7: Classification with Viola&Johnes

Notice that Viola&Johnes classification has small delay, described in 3.4.

actual/predicted	close	medium	wide
close	1841,4	40,5	1858,6
medium	0,0	3624,5	25,5
wide	2997,2	1835,0	17777,3

Table 3: Confusion matrix of first 30s of video [1]

	close	medium	wide
Precision	0,3806	0,6590	0,9042
Recall	0,4923	0,9930	0,7863
Accuracy	0,8260	0,9244	0,7758
F1 score	0,4293	0,7922	0,8411
Specificity	0,8772	0,9127	0,7437

Table 4: Scores of classification of first 30s of video [1]

4.2.2 Optimizer

Optimizer significantly improves the performance of the algorithm as we can see in table 5. It also makes classification more stable. See the difference betwen figure 9 ad 7

	Processed	Total	Ratio	Runtime [s]
with optimizer	235	3455	0,0680174	49,6446
without optimizer	3455	3455	1	601,69

Table 5: Optimizer performance

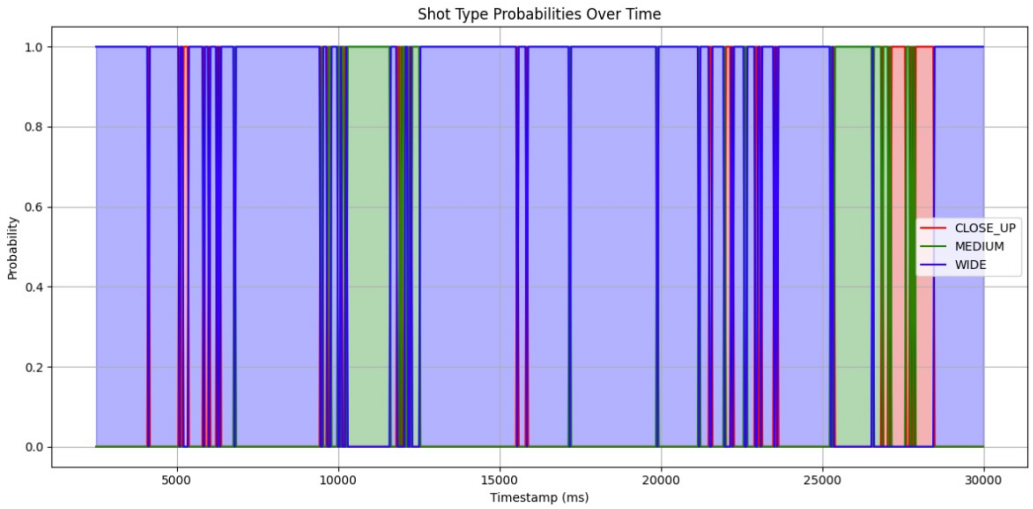


Figure 8: Classification with Viola&Johnes without the optimizer

We adding graph, with oversampling turned on to demonstrate function of this feature and also make data without optimizer more readable.

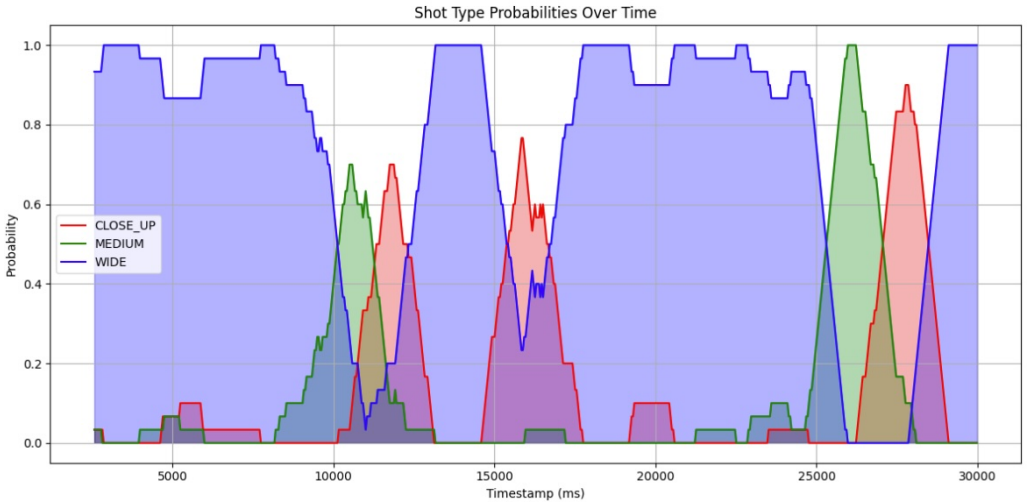


Figure 9: Classification with Viola&Johnes without the optimizer and with oversampling

5 Dataset description

Miroslaw

6 Results summary

Miroslaw

7 Individual contributions

both

8 Conclusion

Marek

References

References

[1] Marvel Studios' Avengers: Infinity War Official Trailer [online]. 2018. [accessed 2025-7-14]. Available at: https://www.youtube.com/watch?v=6ZfuNTqbHE8