# DESCRIPTION FOR THE GENERAL PUBLIC

In the last decade we can observe a quite common presence of software designed for automatic object detection or tracking in digital images or video sequences, e.g.: face detectors (in cameras of mobile devices or Facebook galleries), person or vehicle detectors (in monitoring systems), road signs detectors (in modern cars). The vast majority of those systems is based on an important element called *Haar-like features*, proposed in (Viola and Jones, 2001; Viola and Jones, 2004). They are simple differential features serving as descriptors of analyzed image windows and can be interpreted as rough contours — e.g.: differences between average intensity of forehead and eyes line, nose and cheeks, etc. A suitably large set of such primitive features (e.g. of order $10^4$) may constitute a sufficient basis to distinguish the objects of interest from the background or other objects.

From the algorithmic point of view, the key advantage of Haar-like features is the possibility to extract them very fast, once the so-called *integral image* is prepared. It is an auxiliary array which stores a cumulated function of image intensity and in fact constitutes the essence of the idea. Regardless of the number of pixels in the window under analysis, be it $24 \times 24$ or $256 \times 256$, the number of operations required to extract a single feature remains the same, and such a property in computer science is known as *constant-time* complexity — $O(1)$. This fact is important, because a dense detection procedure, which scans the image with a sliding window, requires a number of calculations for really numerous windows. In general it depends on image resolution and some programistic settings (number of scales, jumps of detection window, etc.), yet, in typical applications the number of analyzed windows for a single image can be of order $10^5$ or even $10^6$. In consequence, the analysis of a single window must be completed within microseconds or faster, so that the whole procedure can be completed in a reasonable time.

In spite of the appealing time efficiency, detectors based on Haar-like features are not faultless. One may observe occasional detection failures or false positive indications. There exist a few modifications of the core idea (e.g. HOG descriptor based on directional gradients or variance/co-variance-based features) which also are constant-time techniques. Nevertheless, they too exhibit some deficiency of accuracy due to the simplicity of mentioned features (their low informativeness). The arguments above were the reasons for choosing this research topic.

The main scientific objective of the project is: *to design new integral representations of images and algorithms based on them, allowing for extraction of features (Fourier moments, Fourier-Mellin moments, statistical moments) from image windows in constant time — $O(1)$ — for the purposes of machine learning and detection tasks*. Therefore, it is authors' ambition to seek new possibilities to extract more advanced features (showing better approximation properties), but still to make the approach based on constant-time complexity.

Apart from mathematical and algorithmic studies, the research to be carried out involves machine learning, in particular: training detectors on large-scale data sets via boosted algorithms, experimentations on different variants of so-called weak classifiers, accelerating techniques ("weight trimming") and regularizations. The obtained results shall be assessed twofold. Firstly, by means of classification accuracy measures (sensitivity, FAR, $F_1$-score, AUC, etc.); secondly, by means of the quality of image reconstructions using the features. Reconstructions can be viewed as approximations of images or image windows. They can be obtained by partial expansions of moment series. Obviously, reconstructions as such are not needed directly for the detection task, but they can be a good gauge of usefulness of the extracted features for detection purposes.