

**Silesian University of Technology**

**Faculty of Automatic Control, Electronics   
and Computer Science**

##### Final Project

##### (choose appropriate)

Implementation of face detection algorithms in video sequences

Author: Wojciech Mazur

Supervisor: PhD Michał Staniszewski

Consultant: PhD Name Surname

Gliwice, January 2019Załącznik Nr 2 do Zarz. Nr 97/08/09

**Oświadczenie**

Wyrażam zgodę/nie wyrażam\* zgody na udostępnienie mojej pracy dyplomowej/rozprawy doktorskiej\*

…………….………., dnia …………………………

……………………………………..……………….……

*(podpis)*

……………………………………………………………

*(poświadczenie wiarygodności podpisu przez Dziekanat)*

*\* właściwe podkreślić*

**Oświadczenie promotora**

Oświadczam, że praca „Implementation of face detection algorithms in video sequences” spełnia wymagania formalne pracy dyplomowej inżynierskiej.

|  |  |
| --- | --- |
| Gliwice, dnia ……………………… | ………………..……………….……  *(podpis)* |

Contents

[1. Introduction 3](#_Toc526891482)

[2. [Problem analysis] 5](#_Toc526891483)

[3. Requirements and tools 13](#_Toc526891484)

[4. External specification 16](#_Toc526891485)

[5. Internal specification 19](#_Toc526891486)

[6. Verification and validation 22](#_Toc526891487)

[7. Conclusions 25](#_Toc526891488)

[Bibliography **Error! Bookmark not defined.**](#_Toc526891489)

[List of abbreviations and symbols iii](#_Toc526891490)

[Contents of attached CD-ROM iv](#_Toc526891491)

[List of Figures v](#_Toc526891492)

[List of Tables vi](#_Toc526891493)

# Introduction

This chapter contains following elements:

* **introduction into the problem domain,**
* **settling of the problem in the domain,**
* **objective of the thesis,**
* **scope of the thesis,**
* short description of chapters,
* clear description of contribution of the thesis’s author – in case of more authors table with enumeration of contribution of authors.

According to Maslow’s Hierarchy of needs, safety is one of our most fundamental needs. Without it, it is hard to think about friends, relationships, accomplishments, or self-fulfillment. Over the last few years it can be observed, that the biological identification is rising in its popularity. There are new ways found to utilize human biological footprints. What can be noticed, is that most of the use cases revolves around security, and for a good reason. Fingerprints are used to unlock locks for decades. Fingerprint scanners have also found a way to consumer electronics. World with securing our smartphone or a notebook with just a password can no longer be imagined. Fingerprint scanners have expended rapidly when it comes to smartphones and made our lives better.

But it’s not the only way people are trying to secure our privacy, including our increasingly valuable resource, which is data. There are ways developed to recognize people by other features like iris, or even from a face in general, using different kinds of sensors. Cameras, infrared light sensors to depth detectors that remember given points on a face. And everything happens within a blink of an eye.

What people want to do apart from protecting our data, is something that humanity crave to do since beginning of itself. To protect ourselves from external harm. Feel safe and comfortable, which is, in fact, required, as stated previously.

One of the main issues arising in recent years, is public safety and detection of dangerous people. Increasing activity of terrorism in a lot of European countries can be observed. The awareness of people is also increasing, it is being learned that prevention is better that treatment. National safety agencies recognize most of European countries at least on moderate threat of terrorism. A lot of countries are considered even as a high threat for regular travelers. This is something that needs to be addressed as quickly as possible, and it is needed to work on the technology, that can be effectively applicated for new solutions.

## Goal of the thesis

Main goal of this thesis is to implement face detection algorithms, both with the face recognition. This could provide us with useful tools to analyze and examine, for example video footages that are collected every day on thousands of city cameras around the globe. That existing architecture can be used, not only to look for people that have already broke the law, but also to prevent crime. Simple camera system doesn’t provide us with the possibility of detecting the face. It would be needed to do that manually which is a complicated and tedious work, that not every person is able to do. The main focus is to accomplish filtering dangerous, or at least for some reason crucial for safety people, implementing algorithms that detect and recognize their faces.

## Scope of work

Main scope of the work contains a few elements that needed to be combined to give full spectrum of the problem. First of all, I needed to choose programming language, that would be easy to use, yet contain all the elements needed to process images, like easily accessible and usable libraries that would greatly accelerate the progress of work. Then, analysis of current solutions, and choice of a few algorithms that comply with what this thesis is trying to achieve and comparison of their effectiveness. Next thing was the actual implementation of the solution in the programming language of choice, but also processing the testing material consisting of video sequences that were analyzed and graphically modified that the human can easily observe effects of the detection and recognition in real time.

# [Problem analysis]

This chapter contains following elements:

* **problem analysis,**
* **state of the art, problem statement,**
* **literature research (all sources in the thesis have to be referenced [1, 2, 4, 3]),**
* **description of existing solutions (also scientific ones, if the problem is scientifically researched), algorithms, location of the thesis in the scientific domain.**

## Face detection methods

There are a lot of ways to detect given features, including face features, that would allow us to recognize position of the face on a given media sample. Although one that got very popular since its release is based on machine learning approach, from the work of P. Viola and M. Jones [1] [2]. Their work is described as extremely rapid in terms of effectiveness, and this is something that is crucial for effective and real time face recognition. The original study was performed on Intel Pentium III clocked on 700MHz, which is a rather old processing unit considering performance of newest units. This would give over 15fps on 384x288 pixel images. Similar CPUs will be compared using tool UserBenchmark [3]. Comparing closest processor that this tool provides us, which is a successor to Pentium III, Pentium 4. Its performance can be taken as comparable, especially in units that were on similar clock speeds [4]. Unit that was used for calculations in this thesis was Intel Core i5 4690k, which is according to UserBenchmark, over ten times faster on average, often going up to twenty times faster and above [3]. This gives us the opportunity to take higher resolution of video samples and photos, and still achieve a very satisfying frames per second rate.

Idea is that if given element, item or other object has a unique shape, it can be assumed, that it also has unique properties when it comes to light that is projected onto that object. This gives us a tool to check whether chosen parts of an image have lighting that is changing in a way that was previously seen and is recognize for example as a face.

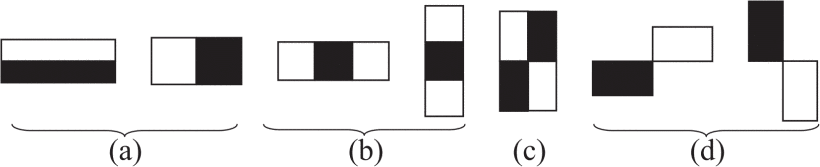


Figure 1Basic Haar-like features [1]

On the Figure 1 is can be seen what features are used to distinguish between light and dark spots on a picture. Although this can be misleading, what should be really be seen if understanding of the problem is a priority, is to see the image in a way shown in Figure 2.



Figure 2Haar-like feature on face [5]

The colors itself are not important, but the pixel values within. For example, if face is to be found, it can be safely assumed, that forehead pixels will be on average brighter than pixels that are within eyes region, due to the shadow dropping from the eyebrows. The same goes mostly for things like nose, where either vertical line, that is bright and has darker surroundings due to the shadows. But this is also something that needs to be taken with care, due to the changing light conditions. Compensation can be done, by looking for white line that has a darker region only on one side.

There were taken many approaches to improve initial Viola and Jones, including the Viola himself. One of the first successful improvements were made only one year later, by the R. Lienhart and J. Maydt [6]. They were dealing with two main problems of original approach, by adding 45 rotated features and adding new optimization procedures for improved performance.

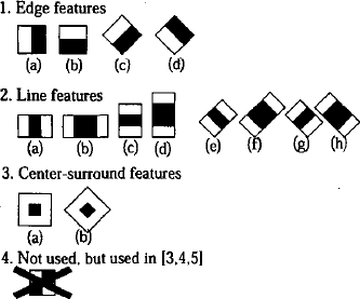


Figure 3Extended features proposed by Lienhart and Maydt [6]

Further attempts of improvements can be observed. T. Mita, T. Kaneko and O. Hori proposed and derived method of detecting co-occurrence of features. One of the things that are addressed by those authors, is that in original solution, after detecting one feature, detecting of proceeding ones comes with much higher error rate [7].

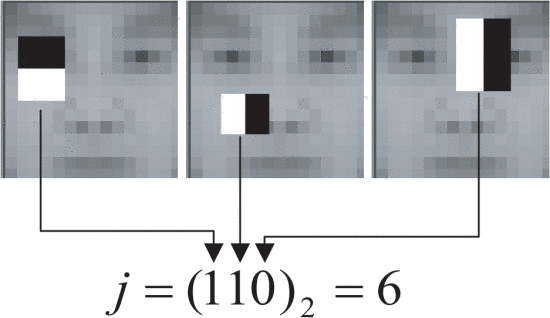


Figure 4 Joint Haar-Like feature

This approach allows to capture more of a structural similarity of faces. Experiment results give reduced error rates and improved performance, even when considering bigger number of features, than in Viola and Jones solution.

## Face recognition methods

Face recognition is a topic that is highly researched for decades, first by manual labor, where for example physical photos were projected onto photomultiplier matrix, then small motors were turned according to illumination [8], which is in fact a machine learning approach. Since then, great amounts of methods were developed to recognize patterns, that includes faces [9]. Numerous approaches that are holistic, hybrid, feature-based, artificial neural networks, fuzzy-based, generic-algorithms based and more can be found [9], where variance in implementation can mixing different approaches can also differ. Some ways of recognition would estimate face position, for instance, to be able to transform image in given space, to obtain more accurate results.

It is also required, for a reasonable efficiency to optimize and normalize dataset from which the algorithm will be built upon. The significant matter is to provide good quality images, for face recognition straight face is a suggested look, without rotation and with the same scale as the rest of the photos, also with an equal lighting [10]. But it is not always possible to provide such resources. Good practice would be to implement image processing that equalize the appearance of the given object.

Next general task in facial recognition is to extract features [9]. Dimensionality of the image is needed to be reduced, so that the data that needs to be processed is not that enormous, and yet the important features are conserved.

### Holistic approaches

Holistic ways of recognizing faces are dealing with the problem with comprehensive approach to processing facial verification, taking data from image as a whole block. They consider small number of features and try to envelop components of an image. Later, these components are used to verify similar shapes on other data [11].

#### Eigenfaces

When considering sets of images, it can be seen that great amount of data can easily arise. Reasonable and natural step would be to reduce its size. The advantage of this approach is that big dataset is not needed, although the bigger is usually the better. Eigenface is a method that is extracting features into vectors, and then represent then in a form of covariance matrix. Having those vectors, it can be calculated how distant they are [12] [11].

However, in this approach training data set is required to be processed, otherwise results will be not satisfactory. It is needed to equalize lighting, and align image, it is also preferred to remove background and other noise that could have negative influence. This algorithm is using PCA (Principal Component Analysis) to reduce dimensions and find vectors. Vector defines subspace – face space, and training set is projected, to find weights, or similarities in other words.

#### Fisherfaces

Fisherfaces are similar to Eigenfaces. Method uses both principal component analysis and linear discriminant analysis and produce a subspace projection matrix. It is minimizing variation within classes and maximizing separation. It should provide better results when it comes to different facial expressions and light variance. Although computation time is more complex, and time needed is greater [13].

### Hybrid approaches

These methods are based on mixing holistic approach and feature-based matching. These two methods have their advantages and disadvantages, but the idea is that combination of those two might give a better result and eliminate each other disadvantages [14] [15].

A lot of hybrid approaches draw from the Gabor wavelets that were acknowledged as a reliable local feature extraction method, due to its sturdiness when it comes to light, distortions, translation [15].

### Feature-based

Every face has some distinctive features related to some regions that can be extracted. Eyes, nose, mouth, cheeks can be used to classify a face. Complexity of these solutions are often hidden under recognizing particular parts of the face. For example, eye processing could be interrupted by reflexes on the iris, that should be eliminated, eyebrows are often described as a parabola approximation. Nose is one of the simplest properties that can be used as a base, where gray levels contrasts are great within neighbor regions. Mouth shape can be also described as a function. Knowing that many facial points, chin location can be also estimated. Approach like this was described in [16] where efficiency was often exceeding 90 percent.

However, this is a limited approach when it comes to providing information about face details. Facial features are rich in texture, but not enough to tell them apart from background. This problem is being approached by adding context information of each feature. But this is a tough task, especially when within- class variance is big.

### Soft computing methods

Face recognition comes with a lot of variance, imprecision, uncertainty and approximations. Soft computing methods are addressing those problems and tries to solve them. These approaches contain methods like fuzzy logic, artificial neural networks, machine learning, generic algorithms [17].

#### Artificial neural networks

Inspired by the biological nervous system and the way it works. This solution aims to solve non-linear problems. Artificial neural networks are in fact an interconnected web of so-called neurons. Each of the neurons performs little operations that adjusts weights, so that desired output is shown. Training is often accomplished by feeding the neural network with patterns

#### Fuzzy logic

This is an approach mimicking human knowledge, that is naturally imprecise. It introduces concept of partial truth and false. That is due to the fact that most of the human body properties are nonlinear and trimming them down to linear solutions makes them often impossible to achieve high accuracy [18]. This method is often incorporated as a part of recognition process with other methods. For example, fuzzy k-nearest neighbor classification to find suitable scatter matrices.

#### Genetic algorithms

Genetic algorithm approaches derive its ide from natural evolution. It is based on inheritance, mutation, natural selection, recombination, where goal is to select fittest solution. It is randomized, but also directed.

Some terms are needed for further explanation to understand basics of those solutions. Chromosomes are set of genes, and gene contains part of the solution. Fitness describes closeness to solution, and its function assign fitness value to individual. Mutation is a change in random gene, crossover is an operation to create new chromosomes for offspring, the better they are, the better chance of “surviving” they have [19].

Solutions can be found where GA solutions are mixed with other approaches, for example systems where principal component analysis is used for feature extraction and Genetic Algorithm to recognize [19].

# Requirements and tools

This chapter contains following elements:

* functional and nonfunctional requirements,
* use cases (UML diagrams),
* description of tools,
* methodology of design and implementation.

## Requirements

### Functional requirements

Basic database of photos should be easy to provide by the user to the given folder. Every photo needs to contain one face, preferably with good lighting. Phots should be grouped in folders, one folder to one person that detection will be take place on. Folders should be named accordingly, uniquely and describe the person, as this name will be used to recognize given face. Every folder should contain the same number of photos, to allow the training process to achieve best performance.

Aligned and processed images should be placed in a separate folder, with the same labels, segregated in folders the same way as before the processing.

The application should provide possibility to specify video sequence that will be further processed as a string of characters.

Face on a processed video media should be tracked and clearly marked with a square shaped box. Above the square should be a text with the prediction result, that is name of the face owner. Text should be readable both when the face detected is far, so the square and text above is small, and when the face is close, and the text could be too big to fit into the screen.

Output video after face detection should be placed in separate folder marked as the output videos, with each video named the same way as before processing with “output” appended to the original title to be easily distinguished.

### Non-functional requirements

## Use cases

## Tools

### Programming language

#### Research results

In data processing, especially in image processing there are a few most viable options when it comes to programming languages. There are basically four different options available.

First option is Java language that comes with two libraries that would help with writing this thesis. Wrapper for OpenCV and native JavaCV. However, due to the lack of experience and personal preferences, choice was made not to use Java for development.

MATLAB is a really powerful tool and can be successfully used for tasks like image processing. There are toolboxes available, that further extend possibilities, for example image processing toolbox, computer vision toolbox and many more. Also, the program allows for easy access to data, and easy visualization, with simple methods. MATLAB also supports OpenCV which is a next huge advantage. It also comes with a great documentation. However, cost of the software is enormous for an ordinary user, that completely brought chances of using this tool to zero percent [20].

More of a low-level approach would be to use C++, it comes with a great support of OpenCV and great performance. Although, it comes with responsibility of managing memory.

#### Language of choice

Language of choice for this project was Python. It is widely used in data science field, due to the huge number of libraries and supportive community. There are a few advantages that decided for this language to be used.

The cleanness of language – this was the main purpose while Python was designed. To keep it simple, consistent and compact, which is already a distinctive feature in a programming world. Code is easy to write and easy to read, and this makes it really great for code maintenance. Yet, it was designed to fully support Object Oriented Programming, imperative, procedural and functional programming.

Efficiency of prototyping – This is a dynamic type language, which means, that there are no strict types defined. Of course, they are assigned later, during interpreting, but a programmer doesn’t have to think about that while creating the code. Python code is often much shorter, which further extends ease of upkeeping clean code.

Portability – Python code is multiplatform for most part. It can be easily executed on many different platforms. It allows not only to make graphical user interface applications, but also web applications. This makes it a universal tool for many applications.

Libraries – Python offers great number of libraries that makes it really easy to develop applications. This includes web development like Django or Flask, calculations like NumPy, game development like Pygame and many more.

Quality of code – On the base of Python stands so-called Zen of Python (PEP 20), that includes aphorisms like “Beautiful is better than ugly” or “complex is better than complicated” or “readability counts”. It has also a lot of other PEP rules defining style convention of Python like PEP 8, where it is defined how many spaces for indentation level, maximum line length or occurrence of white spaces. Even though there are a lot of rules, most of IDE (programming environments) give possibility to install extensions that keep track of those rules and often even automatically correct them. This makes Python code very universal and easy to read among developers.

### Libraries

#### OpenCV

It’s an open source, free for commercial and educational use library. It was written in C and C++ and is developed using wrappers in a lot of programming languages like C#, Python, MATLAB, Java. One of the great advantages of this library possible multiplatform usage. It’s compatible with Windows, Mac OS X, Linux, Android and iOS.

Library was created to serve in real-time applications, where computing efficiency is a main issue. C and C++ as a relatively low-level language gives possibility to optimize and use multicore processor units.

One of the main goals of OpenCV is to provide an all in one tool that would allow to easily create advanced projects by people on different knowledge levels. It can be easily use both for educating beginners, but it also gives the professionals to use advanced tools without reinventing the wheel.

#### NumPy

NumPy package is a basic Python tool that allows for advanced mathematical calculations, mainly for scientific purposes like numerical methods, operations on matrixes, diagonalization, integration, solving of equations and so on.

One of the basic objects that is introduced by NumPy is ndarray. It can be treated as a universal container for data in form of vectors or arrays. The main difference between arrays provided by ndarray and default Python arrays is that objects inside needs to be the same type, they keep their size, while changing size new object is being created, and old is removed, objects contain functions to operate on contained data, optimized to use on big datasets.

#### Matplotlib

Powerful and rich library used for plotting and data visualization. It produces 2D graphics in interactive environments. It allows to draw graphs like lines, scattered, bar and many more.

Matplotlib has Pyplot module that provide MATLAB-like interface. Which makes it easy for developers moving from that environment to begin. It also has object-oriented API to easily embed plotting into graphical interfaces like Qt or Tkinter.

It can be also extended to use with Microsoft Excel, or to create 3D plots by using mplot3d. Active group of developers comes also with great profit with constant support and big community.

#### Scikit-learn

Package that provides simple and efficient tools for data mining and analysis. Includes various classification, regression, clustering algorithms including support vector machines, random forests, gradient boosting, k-means and more. It is also built upon and designed to cooperate with other scientific libraries like NumPy and SciPy.

# External specification

This chapter contains following elements:

* hardware and software requirements,
* installation procedure,
* activation procedure,
* types of users,
* user manual,
* system administration,
* security issues,
* example of usage,
* working scenarios (with screenshots or output files).

### Requirements

|  |
| --- |
|  |
|  |

# Internal specification

This chapter contains following elements:

* concept of the system,
* system architecture,
* description of data structures (and data bases),
* components, modules, libraries, resume of important classes (if used),
* resume of important algorithms (if used),
* details of implementation of selected parts,
* applied design patterns,
* UML diagrams.

A short code insertion in the text line is possible, e.g. class Main. Longer fragments should be written in *Courier* or *Courier New* font size 10 in frames (Listing 4.1) with a space between the lines of the value 1. All lines of code should be numbered so that they can be referenced in the text of the document.

Listing 1. Generating random numbers

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17 | **package** polsl.iinf.lab;  **import** java.util.Random;  **public** **class** Main {  **public** **static** **void** main(String[] args) {  Random r = **new** Random();  // drawing a number from the range 1..10  **int** a = r.nextInt(10 + 1);  System.*out*.println(a);  // drawing a number from the range -5..15  System.*out*.println(r.nextInt(21) - 5);  }  } |

# Verification and validation

This chapter contains following elements:

* testing paradigm (eg. V model),
* test cases, testing scope (full / partial),
* detected and fixed bugs,
* results of experiments (optional).

# Conclusions

This chapter contains following elements:

* achieved results with regard to objectives of the thesis and requirements,
* path of further development (eg. functional extension . . . ),
* encountered difficulties and problems.

# Bibliography

|  |  |
| --- | --- |
| [1] | P. Viola i M. Jones, „Rapid object detection using a boosted cascade of simple features,” w *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, 2001. |
| [2] | P. Viola i M. Jones, „Robust Real-Time Face Detection,” *International Journal of Computer Vision,* tom 57, nr 2, p. 137–154, 2004. |
| [3] | "userbenchmark," [Online]. Available: https://cpu.userbenchmark.com/Compare/Intel-Pentium-4-160GHz-vs-Intel-Core-i5-4690K/m15237vs2432. [Accessed 29 12 2018]. |
| [4] | justinwl, "ancientelectronics," [Online]. Available: https://ancientelectronics.wordpress.com/tag/pentium-4-vs-pentium-iii/. [Accessed 29 12 2018]. |
| [5] | N. Ipe, "N Recursions," 16 10 2018. [Online]. Available: http://nrecursions.blogspot.com/2018/10/a-better-tutorial-on-haar-features-used.html. [Accessed 29 12 2018]. |
| [6] | R. Lienhart i J. Maydt, „An extended set of Haar-like features for rapid object detection,” w *Proceedings. International Conference on Image Processing*, Rochester, 2002. |
| [7] | T. Mita, T. Kaneko i O. Hori, „Joint Haar-like features for face detection,” w *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Beijing, 2005. |
| [8] | W. K. Taylor, „Machine learning and recognition of faces,” IET, London, 1967. |
| [9] | M. &. B. D. Roomi, „A Review Of Face Recognition Methods,” *International Journal of Pattern Recognition and Artificial Intelligence,* tom 27, nr 4, 2013. |
| [10] | T. Shakunaga i K. Shigenari, „Decomposed eigenface for face recognition under various lighting conditions,” w *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, 2001. |
| [11] | G. M. Zafaruddin i H. S. Fadewar, Face recognition: A holistic approach review, Mysore: IEEE, 2014. |
| [12] | P. Belhumeur, J. Hespanha i D. Kriegman, „Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* tom 19, nr 7, pp. 711-720, 1997. |
| [13] | S. Jaiswal, D. (. S. S. Bhadauria i D. R. S. Jadon, „COMPARISON BETWEEN FACE RECOGNITION ALGORITHM-EIGENFACES, FISHERFACES AND ELASTIC BUNCH GRAPH MATCHING,” *Journal of Global Research in Computer Science,* tom 2, nr 7, pp. 187-192, 2011. |
| [14] | R. Huang, V. Pavlovic i D. N. Metaxas, „A hybrid face recognition method using Markov random fields,” w *Proceedings of the 17th International Conference on Pattern Recognition*, Cambridge, 2004. |
| [15] | H. R. R. J. B. C. O. &. M. S. Cho, „An Efficient Hybrid Face Recognition Algorithm Using PCA and GABOR Wavelets,” *International Journal of Advanced Robotic Systems,* tom 11, nr 4, 2014. |
| [16] | P. Campadelli, R. Lanzarotti i C. Savazzi, „A feature-based face recognition system,” w *Proceedings of the 12th International Conference on Image Analysis and Processing* , Milano, 2003. |
| [17] | D. Ibrahim, „An overview of soft computing,” w *12th International Conference on Application of Fuzzy Systems and Soft Computing*, Vienna, 2016. |
| [18] | M. H. Fadzil i L. C. Choon, „Face recognition system based on neural networks and fuzzy logic,” w *Proceedings of International Conference on Neural Networks*, Houston, 1997. |
| [19] | G. M. Zafaruddin i H. S. Fadewar, „Face recognition: A holistic approach review,” w *International Conference on Contemporary Computing and Informatics*, Mysore, 2014. |

# List of abbreviations and symbols

|  |  |
| --- | --- |
| *DNA* | deoxyribonucleic acid |
| *MVC* | model–view–controller |
| *N* | cardinality of data set |

# Contents of attached CD-ROM

The thesis is accompanied by a CD-ROM containing:

* thesis (PDF file),
* source code of applications,

# List of Figures

# List of Tables