Towards Emotion Recognition using Evolutionary Computation

Jing-Ling Tsai

MSc by Research

University of York **ELECTRONICS**

September 2014

Abstract

Facial expression recognition and analysis are difficult because huge amounts of input data needs to be processed and automation is even more complex. However, **Evolutionary Computing (EC)** can be good at complex tasks. For implementation, **Evolvable Hardware (EHW)** is an advantageous technology which is fast in real-time situations. This thesis uses **Evolutionary Computing (EC)** for facial expression analysis, and proposes new algorithms to undertake a limited set of facial expression analysis. Compared with conventional classifiers such as Support Vector Machines (SVMs), results of using a Cartesian Genetic Programming (CGP) classifier show better effectiveness when using the technique. Finally, this thesis discusses **Evolvable Hardware (EHW)** but does not implement it.

Contents

List of Figures

B.983 The 2 run training of 50,000 generations fitness except patterns 2,11 . . . 325

List of Tables

Acknowledgments

I would like to thank my supervisors *Prof. Andy Tyrrell* and *Dr. Steve Smith* who give me guidance and advices in my research. I thank my thesis advisor *Dr. Julian Miller* as well for teaching me Cartesian Genetic Programming (CGP).

I also express my gratitude to my colleague Dr. Martin Trefzer who is so creative in evolvable hardware. The colleagues in the Intelligent System Group in the Electronics Department, especially Pitiwut, Siti, Alfonso, Mike, Luis and Ibrahim, thank you for your accompanying and caring.

Here I would particularly like to express thanks to Holly, Angie, Annette, Maddy and Stuart. I can not carry out my research smoothly without your support and understanding.

Finally, I would like to thank my family and friends who always encourage, support and care me, in particular my mom, my aunt, my cousins, Di-Wen, Rachel, both Ellies, Kate and Hannah.

The experience of studying in UK makes me stronger and better, and those experiences are valuable. Thanks that I have this opportunity to study in York, the charming city in UK!

Declaration

I declare that this thesis is a presentation of original work and I am the sole author.This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1

Introduction

Contents

Machine Learning is the study of algorithms that learn automatically to make predictions based on past data [Schapire, n.d.]. There are many examples of Machine Learning classification such as machine vision for facial recognition, text categorization for email spam detection, and specialized software for financial predictions [Bost et al., 2015] [Schapire, n.d.]. This thesis focuses on discussing the use of Cartesian Genetic Programming (CGP), one type of Machine Learning Algorithm, to implement facial expression classification.

Facial expression recognition has numerous utilizations such as cognitive activity or psychopathology. Facial expressions and gestures communicate non-verbally to the listeners in face-to-face interactions. In some cases, such as identifying criminals, non-verbal communication cues can play an important role because criminals usually tell lies. Faces carry other information to convey expressions, and convey some unwanted information. Automatic systems for facial expressions recognition usually take the form of a sequential configuration of processing blocks, which adheres to a classical pattern recognition model [Chibelushi & Bourel, n.d.]. One approach in automated recognition of facial expression is based on the **Facial Action Coding System (FACS)**, which separates expressions into upper and lower face action [Lien et al., 1998].

Evolutionary Computing (EC) represents a powerful search and optimization paradigm which uses the ideas of natural selection and genetics as biological metaphors [Dumitrescu

et al., 2000*a*]. In the past, **Evolutionary Computing (EC)** has been applied indirectly to face recognition problems. However, **Genetic Programming (GP)** can also be used to develop algorithms that directly discriminate between facial images for a face recognition application [Teller & Veloso, 1995]. For conventional methods of facial expression recognition, human experts design an approach to detect potential image features depending on their personal knowledge and experience. However, **Genetic Programming (GP)** may try many unconventional ways of combining primitive operations that may never be imagined by human experts [Bhanu et al., 2004].

The bio-inspired hardware architecture for high speed image processing classification was proposed in [Glette et al., 2007]. The plan of the project is to help social workers recognise people's emotion and determine whether they are telling the truth or lying from their facial expressions. The datum processing by the hardware is a type of data from software that performs feature extraction and feature recognition. Bitstreams produced from software will be delivered to bio-inspired hardware, then classified from the features of facial expressions illustrated in Figure 1.1. This device helps social workers differentiate not only subjects' expressions but whether they tell truth or not.

Figure 1.1: Project outline of social work

1.1 Motivation

Machine Learning can be seen to be a revolution and has had an impact on Computer Science in the sub-field of Artificial Intelligence. Engineers use computers or instruments to collect huge amounts of data from problems. Afterwards they learn how to resolve those complicated problems by using computers instead of working it out by hand. Genetic Programming is one of the approaches of Machine Learning. It is interesting to apply Cartesian Genetic Programming (CGP) to system on chip (SoC) to form an embedded system device which will process face recognition accurately.

From an application perspective, there are some dark sides to our world. However there are also lots of wonderful things that surround us. Social work is needed to help disadvantaged groups, like abused children, who can be protected by social welfare institutions. In addition to intelligence, devices can be equipped to make life easier, especially when they are specialized, light and portable. They are also winning more and more acceptance. Accordingly, it would be valuable and prospective to do research that combine artificial intelligence and embedded system technologies to assist social work.

It is difficult to distinguish whether people are telling truth or not from their expressions, especially in social work. More tools to aid professions are needed to help make accurate judgements on subjects. In order to achieve this goal, a lie testing device to abstract data from human emotional responses is proposed. The device will help social workers judge whether subjects are abusers or not and then try to rescue victims of abuse. Owning a lie testing device which is reliable and portable for social work could provide a large contribution to this work.

1.2 Hypothesis

" Evolutionary computing is an effective method to distinguish facial expressions. "

The analysis of emotional responses by EHW is a novel field of research. It is a challenge to distinguish which subjects tell the truth from their expressions. We need to cooperate with psychologists to develop training patterns based on the features of telling truth or lies. Following this, one important technology of the system is classification by using evolutionary computing to distinguish facial expressions. Classification accuracy and evolution speed are also factors studied in the experiment. In order to achieve the goals, a system whose behaviours can deal with a complex facial image classification task is developed. The Embedded Cartesian Genetic Programming (ECGP) Architecture [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008; Glette, Torresen, Kaufmann & Platzner, 2008] was proposed and seems a better structure due to the good accuracy of above 90% [Glette, Torresen, Kaufmann & Platzner, 2008]. Simply speaking, CGP has already been an excellent tool to assist a classification task in EHW. The objectives of this work are to give high accuracy, a higher number of inputs and categories, and make it easier to implement in hardware. Having an effective classifier, the high accuracy and speed of classifying features can support excellent results of a testing patterns classification task from subjects. Thus, social workers will have a better device to help in their cases.

1.3 Thesis Organisation

This thesis describes the work of developing a bio-inspired hardware structure implemented to classify images of human emotional responses such as expressions and lie detection from subjects' expressions. Following on from the hypothesis and motivation in Chapter 1, Chapter 2 renders the background. Evolutionary Algorithms (EAs), Cartesian Genetic Programming (CGP), Evolvable Hardware (EHW) and Evolvable Hardware architecure for classification tasks are to be the origin of the research. The Chapter 3 illustrates software simulation using Cartesian Genetic Programming (CGP) to classify 64-bit smile and sad images. After that, the experimental results from Chapter 3 are shown in Chapter 4. Also in Chapter 4, data analysis and an interesting issue which is pointing out and doing validation of the best chromosome of the experiment from Chapter 3. Chapter 5 proposes the implementation of Evolvable Hardware (EHW) which are described in the first two sections, and section 5.3 proposes the idea of hardware implementation referred to in Chapter 3. Finally, conclusions and future work are exhibited in Chapter 6.

Chapter 2

Evolutionary Computing

Contents

Evolutionary Computing expresses a powerful search and optimization pattern [Dumitrescu et al., 2000*a*]. The search space is composed of all possible solutions of the problem. The problem of improving parameter values for an application is then transformed into the problem of searching for better solutions in the solution space [Dawkins & Bentley, 1999].

The history of **Evolutionary Computing** goes back to the nineteen forties. As early as 1948, "genetical or evolutionary search" was proposed by Turing, and by 1962 Bremermann had executed computer experiments on "optimization through evolution and recombination" [Eiben & Smith, 2003*a*]. During the 1960, there were three different implementations of the fundamental idea developed. These where : (1) Fogel, Owens, and Walsh introduced **Evolutionary Programming (EP)** in the USA [Fogel et al., 1966], (2) Rechenberg and Schwefel invented **Evolution Strategies (ES)** in Germany [Rechenberg, 1973] [Schwefel, 1995], and (3) a **Genetic Algorithm (GA)** was created by Holland [Holland, 1973] [Holland, 1992*a*]. These three fields developed respectively for almost 15 years, but since the early 1990s they have been viewed as different representations of one technology which has been known as **Evolutionary Computing**. In the early 1990s, **Genetic Programming (GP)** was the forth area following the general idea being created by Koza [Koza, 1992*a*] [Koza, 1994]. The modern terminology of **Evolutionary computing** covers all of these fields, and the algorithms are termed **Evolutionary Algorithms (EAs)** [Eiben & Smith, 2003*a*]. [Martin, 2003] describes the history of Evolutionary Computing. Figure 2.1 shows the phylogeny of Genetic Programming (GP) which is the same as the phylogeny of the Evolutionary Algorithm, take from figure 2.1 of Peter Martin's thesis.

Figure 2.1: The phylogeny of GP as a timeline [Martin, 2003]

This chapter renders background of the Evolutionary Algorithm (EA), including what the Evolutionary Algorithm (EA) is and details Cartesian Genetic Programming (CGP). Furthermore, the last two sections introduce research on Evolvable Hardware (EHW) and previous research corresponding to implementing classification tasks by Evolvable Hardware (EHW).

2.1 Principles of the Evolutionary Algorithm (EA)

In order to resolve tough problems, measures have been developed which mimic natural evolution and self-adaptation. The computer algorithms that are in accordance with these functions are called **Evolutionary Algorithms (EAs)** [Greenwood & Tyrrell, 2007*a*]. They have been developed through stochastic and optimization methods via experimental solutions. The main characteristic of Evolutionary Algorithms (EAs) is the intensive use of randomness and genetics-inspired operations to evolve a set of candidate solutions. The candidate solutions to a certain problem are knowns as **chromosomes** or **individuals**. A chromosome is composed of **genes**. Each gene is located at a particular position on the chromosome called the **locus**. Furthermore, a gene may takes several values - or may have several forms - , each value of which is known as an **allele** [Dumitrescu et al., 2000*a*].

According to natural selection which exhibits survival of the fittest population, figure 2.2 illustrates a heuristic estimation solution of an Evolutionary Algorithm (EA) in a flow chart. First of all, genes are encoded compose chromosomes and generate an initial population randomly. Evaluating each candidate according to a fitness function produces a set of new candidates. On the basis of the fitness function, high scoring candidates are chosen to become parents. Variations are then produced through recombination and/or mutation. The candidates produced from variation being offsprings by selection turn into the new parents for creating the next generation. If the solution satisfies a predefined ending standard, it will be exported and the experiment terminated. If not, the evolution will recur and the solution will be evaluated again. This process can be iterated until the terminal criteria is achieved. In addition, sometimes it will have to go back to initialization from variation or selecting more data if the fittest candidates are not chosen.

Figure 2.2: The general scheme of an EA

2.1.1 Elements of an Evolutionary Algorithm (EA)

In order to define a particular Evolutionary Algorithm (EA), a number of elements, procedures, and operators of Evolutionary Algorithms (EAs) are specified [Eiben & Smith, 2003*b*]. The elements of Evolutionary Algorithms (EAs) indicated in figure 2.2 are defined. Referring to systematization in [Greenwood & Tyrrell, 2007*b*], they are :

- *•* Representation
- *•* Population
- *•* Evaluation(fitness function)
- *•* Selection
- Variation(recombination and mutation)
- *•* Terminal Criteria

Each item is explained as below. Furthermore, for the purpose of making the algorithms run smoothly, an initialization procedure and terminal conditions are also defined.

Representation

Linking the real world and the Evolutionary Algorithm (EA) world, sets up a bridge between the original problem context and the evolutionary space and is the first step in definition of a Evolutionary Algorithm (EA) in which evolution takes place. This first design step is commonly called **representation**. In biological term, it often adopts **phenotype** and **genotype** to describe representations. The possible solutions within the problem context are referred to as **phenotypes**. After their encoding, **genotypes** are formed from the **phenotypes** [Eiben & Smith, 2003*b*]. For instance, when utilized in Evolvable Hardware (EHW) problems, the genotype is the subset of all possible encoded problem parameters to describe a particular circuit configuration. The phenotype is the actual problem solution [Greenwood & Tyrrell, 2007*b*]. In addition, "representation" can also be used in a different sense, which is not focused on mapping, but on the "data structure" of the genotype space [Eiben & Smith, 2003*b*].

What follows here describes an example of the **binary strings** representations from [Greenwood & Tyrrell, 2007*b*] which are arguably the most frequently encountered encodings. Consider a *RC* network which can be encoded with a 6-bit binary string. Possible resistor values are *R* = *{*100Ω*,* 220Ω*,* 390Ω*,* 470Ω*,* 1*K*Ω*,* 2*.*2*K*Ω*,* 3*.*3*K*Ω*,* 5*.*6*K*Ω*}* and possible capacitor values are $C = \{10\mu F, 47\mu F, 220\mu F, 1000\mu F\}$. Designers can use three bits to select one of eight *R* values such as 000 for 100 Ω , 001 for 220 Ω and so on. Also, two bits are selected to represent one of four *C* values such as 00 for 10 μ *F*, 01 for 47 μ *F* and so on. Finally, one bit is used to indicate whether the *R* and *C* are in series (logic 0) or parallel (logic 1). For example, the binary string 011101 would indicate a 470 Ω resistor in parallel with a 220 μ F capacitor.

Population

The **population** is a set of individuals (also called chromosomes or candidate solutions) that consists of bit strings or numbers as shown in figure 2.3.

Figure 2.3: The directions of population

The role of the population is to store possible solutions. A population is a multiset of genotypes because the population may involve more than one copy of the same individual [Eiben & Smith, 2003*b*] [Greenwood & Tyrrell, 2007*b*]. The initial population is usually generated randomly, but may include illegal solutions. Referring to the suggestion of [Greenwood & Tyrrell, 2007*b*], designers could create one legal solution by hand, then duplicate it *N −* 1 times for a population with *N* individuals. Following this, performing the variation operator on this population would create a random population made up entirely of strictly legal individuals. The population size also influences solutions. If the size is large, it will possibly obtain the best solution but will result in high computing consumption. Otherwise, the solution might be locally optimal and may not achieve a global solution.

Evaluation(fitness function)

Before describing the fitness function, Greenwood and Tyrrell focussed primarily on the **objective function** in their book. Evaluation indicates the quality of population through a numeric score to a solution called the **objective function**. The definition of this function is rendered as follow : The objective function maps a solution onto the real number line. If *O* is an objective function and *X* is the space of all possible solutions, then $O: X \to \mathbb{R}$. Solution *α* is better than solution β if $O(\alpha) > O(\beta)$.

The evaluation process assign a fitness value to each chromosome or individual sampled by the evolutionary algorithm. In nature, the fitness of an individual means how well the organism adapts in a particular environment. Comparably, according to a particular problem specification, fitness provides a measure of how well an individual or a solution performs in the evolutionary algorithm [Vellasco et al., 2001]. A **fitness function** is a particular type of objective function that is used to determine the ability of solutions to adapt to their environment [MIBI, n.d.]. Evaluating the parameter sets and choosing which parameter sets mate is based upon the survival of the fittest individuals. Therefore, an appropriate fitness function can distinguish between good or bad chromosomes, and the values can be minimized or maximized depending on the definition.

The fitness value is associated with each point in the search space. If there is only one objective to be satisfied in the particular search problem, the fitness function will usually be straightforward. However, it is complex when more than one objective must satisfied by the evolutionary algorithm [Vellasco et al., 2001]. This issue will be discussed in chapter 2.1.2.

Selection

The role of **selection** is to distinguish between individuals based on their quality. It often depends on an individual's fitness. After initialization, designers usually use selection to allow the better individuals to become parents of the next generation. Selection of a parent is responsible for further quality improvements. The surviving selection undergoes variation after creating the offspring of the selected parents. This is also based on their higher quality fitness values. Different from parent selection that is typically stochastic, survivor selection is often deterministic, which is usually chosen from selections with the top fitness values [Eiben & Smith, 2003*c*].

Two selection methods obtained from [Greenwood & Tyrrell, 2007*b*] chosen to do selection are named **Truncation Selection** and **Tournament Selection**. The **Truncation Selection** method selects individuals from the previous population to construct the next population. For example, the μ parents from $P(t)$ undergo variation to produce λ offspring. All $\mu + \lambda$ are collected and sorted by fitness values. The μ best fit individuals are retained to form $P(t+1)$ and the rest are discarded. **Tournament Selection** takes a random uniform sample of $q > 1$ individuals from the population and then selects the highest fit individual.

Tournament Selection method is stochastic, which means that if the globally optimum solution is found, it could be accidentally lost. One way to prevent this is by using **elitism** where a small number of the best fit individuals are copied unchanged from $P(t)$ to $P(t+1)$. **Truncation Selection** is naturally elitist.

Variation(recombination and mutation)

The variation operator creates new individuals from existing ones by changing some or all of the encoded solution parameters [Greenwood & Tyrrell, 2007*b*]. There are two types of variation operator which are **mutation** and **recombination** illustrated as follow.

• Mutation

The idea of mutation is to create a single offspring by randomly altering encoded solution parameters in the parents. This can result in new gene values which an EA may be able to use arrive at a better solution than ones previously possible. There are two examples encoded as an integer array shown as figure 2.4. The top example is of two randomly chosen integers being swapped. The bottom of figure 2.4 shows another mutation method which reverses the order of all integers between two randomly chosen locations. This method is known as inversion [Greenwood & Tyrrell, 2007*b*].

Mutation

ex: (a) Swapped :

parents offspring	125873640 124873650
(b) Inversion :	
parents offspring	125873640 125637840

Figure 2.4: Mutation [Greenwood & Tyrrell, 2007*b*]

The mutation rate is usually low and the best value is problem dependent. The performance of EA will degrade if the mutation rate is too low. On the contrary, the evolutionary process will approach a random search if the mutation rate is too high. Generally speaking, the value usually recommended is between 1% and 5% [Vellasco et al., 2001].

• Recombination

Another variation operator uses genetic material from two or more parents to generate offspring. This is called **recombination** or **crossover**. It is the process of choosing more than one parent solution to produce a child solution. Figure 2.5 explains three methods of crossover. 1-point crossover is the simplest recombination operator in which bits are exchanged after the crossover point to produce offspring. The middle method of the figure shows 2-point crossover. The bits between the two crossover points, with the others staying the same, are swapped from the parents to become children. The last method is uniform crossover in which one offspring is created from two parents by treating each bit independently. Each bit has an associated random variable. If the random variable value is greater than 0.5, the bit from the upper parent is copied to the offspring. Otherwise the bit from the lower parent is chosen [Greenwood & Tyrrell, 2007*b*].

Recombination

Figure 2.5: Recombination [Greenwood & Tyrrell, 2007*b*]

One-point and two-point crossover have a local scope, and uniform crossover can combine arbitrary patterns present in the chromosome. Furthermore, uniform crossover is also more likely to destroy good features of the individuals. There are also some documents showing crossover involving more than two individuals [Vellasco et al., 2001].

The variation operators are representation dependent. This means different representations have to define different variation operators. For instance, when genotypes are bit-strings, a mutation operator inverts a 0 to a 1. However, if the representations are tree-like structures, another mutation operator will be required [Eiben & Smith, 2003*b*].

Terminal Criteria

EAs are stochastic and there is mostly no guarantee that they will arrive an optimal solution, so the condition might never be satisfied or the algorithms also might never stop [Eiben & Smith, 2003*b*]. Referring to [Greenwood & Tyrrell, 2007*b*], three criteria are commonly used to decide whether to stop the algorithms which are described as the following :

- The algorithm has converged
- *•* A fixed number of generations have been processed
- A sufficiently good enough solution has been found

These three terminal criteria seem to contain all the possible stop conditions which are optimal values achieved or conditions satisfied. Except from the solutions reach the fittest value of objective function, the Evolutionary Algorithm (EA) also needs other end requirements to prevent unlimited runs.

2.1.2 Multiobjective Evolutionary Algorithms

Following on from the Evaluation section, a single-objective function is well-suited to define standard Evolutionary Algorithms (EAs) for optimizing problems. However, most real-world applications tend to require simultaneous optimization for more than one objective [Jong, 2006*a*]. One simple answer is to convert the multiple objectives into a single objective function. For example : a weighted sum $F(x) = \omega_1 F_1(x) + \omega_2 F_2(x) + \cdots + \omega_n F_n(x)$. The challenging issue of this case is the fact that these weights are difficult or impossible to obtain. In this case, instead of "guesstimating" the unknown weights, it is better to provide a set of "tradeoff" solutions. Also, the particular solutions can be selected according to other external factors [Jong, 2006*b*]. Consider the design of a house. The ideal design could maximise space and quality of location while minimising cost. Obviously, these objectives are conflicting. More space in a house may have high cost, and a low-cost house might be in a worse location. Hence, there is no single optimal objective to be found. Multiobjective Evolutionary Algorithms (MOEAs) was proposed to generate all optimal sets of solutions which give designers evaluating the trade-off options between different choices. This could be choosing between two different houses in accordance with different marketing needs: a narrower house means less expense, alternatively a high-cost house which possesses a highvalue location may be the major objective. In addition, Evolutionary algorithms are suitable for using multiobjective optimization because they are based on biological processes being naturally multiobjective [Müller, 2002].

In the mid-1980's, the first pioneering work in solving multiobjective optimization problems using evolutionary approaches was created by Schaffer and Fourman [Zitzler & Thiele, 1999] [Zitzler Eckart et al., 2000]. This led to a number of algorithms being successfully proposed and applied in various problems [Zitzler Eckart et al., 2000] [Zitzler et al., 2001] [Zitzler & Thiele, 1999].

Definations [Zhou et al., 2011] [Zitzler Eckart et al., 2000] [Müller, 2002]

In traditional optimization problems, multiple and conflicting objectives are aggregated into a single-objective optimization problem. The difficulties are mentioned above. Here, a set of optimal trade-offs are found which is the so-called **Pareto-optimal front**.

In contrast to fully ordered scalar search spaces, multiobjective search spaces are partially ordered such that two arbitrary solutions are related to each other in two possible ways: either one dominates the other or neither of them are dominant.

A multiobjective optimization problem can be formulated mathematically as :

$$
minimize F(x) = (f_1(x), \cdots, f_m(x))^T
$$
\n(2.1)

 $s.t. x \in \Omega$

where Ω is the decision space and $x \in \Omega$ is a decision vector. $F(x)$ consists of *m* objective functions $f_i: \Omega \longrightarrow \mathbb{R}, i = 1, \dots, m$, where \mathbb{R}^m is the objective space.

The objectives in formula (2.1) usually conflict with each other such that if one is improved, it may result in degeneration of the others. Therefore, a single solution to optimize all objectives simultaneously is impossible. Rather, it is required to estimate the **Pareto optimal** solutions to give the best trade-off solution. Edgeworth and Pareto primarily proposed the **Pareto-** **Optimality**, for which the formal definitions can be found in [Zitzler Eckart et al., 2000], [Müller, 2002] and [Zitzler et al., 2003].

Definition 1 :

A vector $u = (u_1, \dots, u_m)^T$ is represented to dominate another vector $v = (v_1, \dots, v_m)^T$, denoted as $u \prec v$, if and only if $\forall i \in \{1, \dots, m\}$, $u_i \leq v_i$ and $u \neq v$.

Definition 2 :

A feasible solution or non-dominated solution [Eiben & Smith, 2003*d*] *x ∗ ∈* Ω of formula (2.1) is named a **Pareto optimal solution**, if and only if $\exists y \in \Omega$ such that $F(y) \prec F(x^*)$. The set of all the Pareto optimal solutions is called the **Pareto set (PS)** which is denoted as : $PS = \{x \in \Omega \mid \exists y \in \Omega, F(y) \prec F(x)\}$

The image of the Pareto set (PS) in the objective space is called the **Pareto front (PF)** $PF = \{F(x) \mid x \in PS\}$

Applications

There is a large volume of literature illustrating distinct categories of Multiobjective Evolutionary Algorithms (MOEAs) such as the Vector Evaluated Genetic Algorithm (VEGA) created by Schaffer in 1984, the Nondominated Sorting Genetic Algorithm (NSGA) produced by Srinivas and Deb in 1994, and most recently the Strength Pareto Evolutionary Algorithm (SPEA) proposed by Zitzler and Thiele in 1999. [Zitzler et al., 2003] and [Zhou et al., 2011] explain Multiobjective Evolutionary Algorithms (MOEAs) problems while different algorithms are compared and applied in [Zitzler & Thiele, 1999] and [Zitzler Eckart et al., 2000].

Corresponding to the nature of biological processes in the real world, multi-objective instances are more reasonable than single-objective cases. This applies to solving optimization problems. Various fields of scientists and engineers use Multiobjective Evolutionary Algorithms (MOEAs) as they are effective and robust. [Zhou et al., 2011] list the diverse areas and applications of MOEAs such as "an elitist multiobjective genetic algorithm (EMOGA)" used in pattern classification, "a novel multiobjective optimization immune algorithm in dynamic environment" for greenhouse control, "a multiobjective optimization approach based on prototype optimization with evolved improvement steps" applied in financial optimization, etc.

The purpose of Multiobjective Evolutionary Algorithms (MOEAs) is to approach a set of Pareto optimal solution rather than a single solution. It is worth probing how to generate offspring efficiently. Another interesting view found in [Zhou et al., 2011], describes interactive Multiobjective Evolutionary Algorithms (MOEAs), where interaction with decision makers involved. This paper also explains how to model or classify noises and dynamic environments when using multiobjective optimisation and also paves the way for future work.

2.1.3 Schema Theory

When considering Evolutionary Algorithms (EAs) some questions may come naturally : Why should Evolutionary Algorithms (EAs) build solutions of better and better fitness quality? Why should they find a satisfactory solution for a given problem? What is the probability of improving the fitness quality for solutions with each generation? What is the probability of finding a satisfactory solution to a given problem? In order to answer theses questions, it is possible by means of precise mathematical models to answer these questions. The most studied and best known classes of models in Evolutionary Algorithms (EAs) are commonly known as **Schema Theories** [Cagnoni & Vanneschi, 2010].

The **Schema Theory** was put forward in Holland's book [Holland, 1992*b*] (first edition in 1975, The University of Michigan). Most description of the theory in this section comes from Holland's book. In an adaptive system, the major challenge is that the set of possible structures α is very large and the performance functions μ_E comprise many local maxima. Designers could proceed the search of α if there will be significant improvements. However, unexploited probabilities may doom the system to a fruitless search. Therefore, consistently testing and incorporating structural properties associated with better performance, as an integral part of the search of α , is only one insurance of these contingencies. In view of this, the useful properties have to be determined and exploited. This suggests that the useful properties should also be points of comparison between structures that yield better-thanaverage performance.

Let us assume a finite set of detectors $\{\delta_i : \mathfrak{a} \to V_i, i = 1, \dots, l\}$ are given. According to the detectors, each structure $A \in \mathfrak{a}$ will have a representation $(\delta_1(A), \delta_2(A), \cdots, \delta_l(A)).$ This means that the detector observes values $\delta_i(A) \in V_i, i = 1, \dots, l$ for each structure A. Let the symbol " \Box " mean " don't care ". $(v_{13}, \Box, \Box, \cdots, \Box)$ designates the subset of all elements in **a** having the attribute $v_{13} \in V_1$. Equivalently, $(v_{13}, \Box, \cdots, \Box)$ designates the set of all *l*-tuples in a beginning with the symbol v_{13} . For instance, if $l = 3$, (v_{13}, v_{22}, v_{32}) and (v_{13}, v_{21}, v_{31}) belong to (v_{13}, \Box, \Box) , (v_{12}, v_{22}, v_{32}) will not. [Holland, 1992*b*]

The set of all *l*-tuples including "don't care" and attributes is provided by the augmented

product set $\Xi = \prod_{i=1}^{l} \{V_i \bigcup \{\Box\}\}\.$ Any *l*-tuple $\xi = (\Delta_{i_1}, \Delta_{i_2}, \cdots, \Delta_{i_l}) \in \Xi$ designates a subset of $\mathfrak a$ as following : $A \in \mathfrak a$ belongs to the subset if and only if (1) when $\Delta_{i_j} = \Box$, any attribute from V_j may happen at the *j*th position of *A*, (2) when $\Delta_{i_j} \in V_j$, the attributes Δ_{i_j} must happen at the *j*th position of *A*. For example, $(v_{11}, v_{21}, v_{31}, v_{43})$ and $(v_{13}, v_{21}, v_{32}, v_{43})$ belong $\mathcal{L}(\Box, v_{21}, \Box, v_{43})$ but $(v_{11}, v_{21}, v_{31}, v_{42})$ does not. The set of *l*-tuples $\xi = (\Delta_{i_1}, \Delta_{i_2}, \cdots, \Delta_{i_l})$ belonging to Ξ will be named the set of **schemata**. [Holland, 1992*b*]

The Schema Theory for Genetic Algorithm (GA)

The Schema Theory is applied to the Standard Genetic Algorithm (SGA) using fitness proportional selection, one-point crossover, and bitwise mutation [Eiben & Smith, 2003*e*]. [Malcolm Heywood, 2012], [Eiben & Smith, 2003*e*] and [Dumitrescu et al., 2000*b*] explain this issue, firstly by defining three things : (1) an example of a schema *H*, (2) a schema of order $o(H)$ which means the number of positions in the schema that do not have the " \square " sign, and (3) a schema defining length $d(H)$ which is the distance between the first and last non " \square " gene in schema *H* (this equals the number of possible crossover points between these positions). Thus, the schema $H = 1\square \square \square \square \square 10\square$ has order $o(H) = 4$ and defining length $d(H) = 7 - 1 = 6$. "□" represents "don't care".

What follows is an attempt to model the probability $P(H, t)$ of the schema *H* being selected at generation *t*. Consider the following two parts of models :

- *•* Probability of selection *P*(*H, t*) *∝* number of individuals matching schema *H* at generation *t* denoted as $m(H, t)$;
- *•* Probability of selection *P*(*H, t*) *∝* average fitness of individuals matching schema *H* relative to the average fitness of all individuals in the population. The former average fitness is denoted as $f(H, t)$, and the latter as $\bar{f}(t)$.

Therefore, depending on the models above, a formula is obtained,

$$
P(H,t) = \frac{m(H,t)f(H,t)}{M\bar{f}(t)}
$$
\n(2.2)

Where *M* indicates population size, or the number of independent samples taken to create the next set of parents. Under fitness proportional selection, the expected number of instances

of schema *H* in the population is,

$$
E[m(H,t)] = M \times P(H,t) = \frac{m(H,t)f(H,t)}{\bar{f}(t)}
$$
\n(2.3)

Next, search operators, one-point crossover and bitwise mutation, are introduced which may cause the schema to be disrupted. Consider a genotype of length *l* that contains the schema *H*. Under one-point crossover, the (lower bound) probability of schema *H* surviving at generation *t* is,

$$
P(H \text{ survives}) = 1 - P(H \text{ does not survive}) = 1 - \frac{d(H)}{l-1} p_{diff}(H, t)
$$
\n(2.4)

Where $p_{diff}(H, t)$ is the probability that the second parent does not match schema *H*. In addition, $p_{diff}(H, t) = 1$ is the special example of the worst case lower bound. The other search operator is bitwise mutation. In order for schema *H* to survive, all non " \square " genes in the schema remain unchanged as much as possible. Accordingly, the probability that a certain position of the schema *H* will survive the action of mutation is 1*−pm*, where *p^m* is the mutation probability that a certain position will change its value. The schema *H* will survive the application of the mutation operator whose probability is $(1 - p_m)^{o(H)}$. Typically, if the mutation probability p_m is small $(p_m \ll 1)$, $(1 - p_m)^{o(H)} \approx 1 - o(H)p_m$ may be obtained.

After normalising by *M* to make the result population-size independent, allowing for the disruptive effects of recombination and mutation, and using an inequality to allow for the creation of new instances of *H* by the variation operators, the expected number of schema *H* at generation $t + 1$ when applying to a standard genetic algorithm (SGA) with proportional selection, one-point crossover, and bitwise mutation is,

$$
E[m(H, t+1)] \ge m(H, t) \frac{f(H, t)}{\bar{f}(t)} (1 - p_c \frac{d(H)}{l - 1} p_{diff}(H, t)) (1 - o(H)p_m)
$$
 (2.5)

The parameters p_c and p_m are the rate of crossover and mutation, respectively. If we disregard the small product term $(p_c \frac{d(H)}{d_{c-1}})$ $\frac{d^{(1)}(H,t)}{d-1}$ *pdiff* (H,t) *o*(*H*)*pm*) from the formula (2.5), we obtain the fundamental theory of genetic algorithms to be the following :

$$
E[m(H, t+1)] \ge m(H, t) \frac{f(H, t)}{\bar{f}(t)} (1 - p_c \frac{d(H)}{l - 1} p_{diff}(H, t) - o(H)p_m)
$$
 (2.6)

This is the **schema theory**, and the form is naturally specific under the selection and search operators from which it was derived. A more generic form for the schema theory might be represented as :

$$
E[m(H, t+1)] \ge m(H, t)\alpha(H, t)(1 - \beta(H, t))
$$
\n(2.7)

Where $\alpha(H, t)$ is the "selection coefficient" and $\beta(H, t)$ is the "transcription error". Specifically, for schema *H* survive, then,

$$
\alpha(H,t) \ge (1 - \beta(H,t)) \Longrightarrow \frac{f(H,t)}{\bar{f}(t)} \ge (1 - p_c \frac{d(H)}{l-1} p_{diff}(H,t) - o(H)p_m) \tag{2.8}
$$

Summary

Holland's schema theory analyses the expected schema sampling behaviour of next generations applied to Genetic Algorithms (GAs). The result obtained indicates that the frequency of schemata with above-average fitness, which means short schemata of low order, tends to increase in the next generation. Schemata possessing short definition length, low order, and high performance are known as **building blocks**. However, the schema theory provides only lower bounds for the expected number of instances of a schema in the next generation. The **building block hypothesis** suggests that Genetic Algorithms (GAs) are able to evolve good solutions by combining these fit, low order, short definition lengths schemata to form better solutions. For example, for a population made up of strings of length $l = 7$, the following schemata may represent building blocks: $H_1 = 10$ □□□□, $H_2 = \Box$ □11□□, and $H_3 = \Box \Box \Box \Box \Box$ Due to the action of Genetic Algorithms (GAs), building blocks combine in order to form bigger and bigger building blocks with better and better performances [Sekanina, 2004*a*] [Dumitrescu et al., 2000*b*].

2.1.4 Types of Evolutionary Algorithm (EA)

The Evolutionary Algorithm (EA) is a popular term and is generally divided into three approaches which are **Evolutionary Programming (EP)**, **Genetic Algorithms (GA)**, and **Evolution Strategies (ES)**. Today designers tend to blend these algorithms leading to the distinctions between them becoming blurred [Greenwood & Tyrrell, 2007*c*]. **Genetic Programming (GP)** is somewhat different from Evolutionary Algorithms (EAs) which are a specialised form of Genetic Algorithm (GA). In this subsection, ES, GA and GP are introduced.

Evolution Strategy (ES)

The Evolution Strategy (ES) works with parameter control via self-adaptation [Hansen et al., 2013]. The Evolution Strategy (ES) usually works with small populations and does not use recombination, which indicates that it is just mutation driven [Miller & Tempesti, 2011]. Selection in evolution strategies is based on fitness rankings, not on actual fitness values. The conceptual algorithm of the self-adaptation evolution strategy is described below [Beyer, 2007]:

- 1. Initialize parent population $P_{\mu} = \{a_1, \dots, a_{\mu}\}.$
- 2. Generate the offspring population by mutating the parent $Q_{\lambda} = \{b_1, \dots, b_{\lambda}\}.$
- 3. Select new parent population by deterministic truncation selection from either
	- the offspring population Q_λ : which means that the new population is generated from the best individuals of the offspring only, meaning that there is only offspring competition, denoted as (μ, λ) , or
	- the parent P_μ and offspring Q_λ population: which means that the new population is generated from the best individuals of parents and offspring, meaning that is parent and offspring compete with each other, denoted as $(\mu + \lambda)$.
- 4. Goto step 2 until terminal criterion fulfilled.

Genetic Algorithm (GA)

A Genetic Algorithm (GA) is a search algorithm based on the mechanics of natural genetics and selection [Goldberg, 1989]. The evolutionary process is similar to figure 2.1. A population of strings called chromosomes, which are encoded candidate solutions of named individuals to an optimization problem, evolves towards better solutions. An algorithm is initialized with a population represented by chromosomes. Solutions from one population are taken to generate a new population called the offspring, which are selected according to their fitness. After variation, the more suitable the candidates are, the more opportunities they obtain to be produced. This is motivated by a hope that the new population will be better than the previous one. The process is repeated until the terminal criterion are satisfied [Obitko, 1998].

There are three aspects of applying Genetic Algorithms (GAs) [Wall, n.d.] [Mitchell, 1998] :

• definition of the fitness function: Individual solutions are selected through a fitness based process where fitter solutions measured by a fitness function are typically more likely to be chosen. A simple example would be maximising the real-valued onedimensional function

$$
f(y) = y + |\sin(32y)|, \ 0 \le y < \pi \tag{2.9}
$$

The candidate solutions are the values of *y*, and they can be encoded as bit strings representing real numbers. The fitness calculation translates a given bit string *x* into a real number *y*. After this, *y* is used to evaluate the function $f(y)$. The fitness of a string is the function value at that point.

- definition and implementation of the genetic representation: One approach of the genetic representation is to encode solutions as binary strings where the digit at each position represents the value of some aspect of the solution such as *{*1011001011000010*}*. Another similar approach is to encode solutions as arrays of integers or decimal numbers, with each position representing some particular aspects of the solution like *{*1*,* 5*,* 3*,* 2*,* 6*,* 4*,* 1*,* 4*,* 9*}*.
- definition and implementation of the genetic operators: This step generates a next generation population of solutions from those selected through the genetic operators crossover (also called recombination) and/or mutation, which influence the performance.

Once the three conditions have been well defined, the genetic algorithm should work well. Moreover, in order to improve performance, designers can try some different variations such as using multiple optima or parallelizing the algorithms [Wall, n.d.].

Genetic Programming (GP)

Genetic Programming (GP) continues the trend of dealing with the problem of representation. There are many problems in representation with fixed-length character strings. For example, the most natural representation for a solution is a hierarchical computer program. Because the shape and the size of the hierarchical computer program will solve a given problem which is generally not known beforehand, the program should have the potential to change the size and shape. Genetic Programming (GP) operates very general and hierarchical computer programs [Koza, 1992*b*].
Figure 2.6 shows an example of Genetic Programming (GP) representation, which depicts graphically the LISP S-expression as a rooted, point-labeled tree with ordered branches.

Figure 2.6: An example of GP representation

The four internal points of the tree are labeled with functions (*∗, ∗,* +*, and∗*), and the five external points (leaves) are labeled with terminals (*Z, Y, Y,* 0*.*317*, andZ*). So the S-expression of the tree is *ZY* (*Y* + 0*.*317*Z*) [Koza, 1992*b*]. The evolutionary process of Genetic Programming (GP) is similar to figure 2.1 because it is also a type of Evolutionary Algorithm (EA).

2.2 Cartesian Genetic Programming (CGP)

Cartesian Genetic Programming (CGP) originates from a method of evolving digital circuits developed by Julian Miller in 1997. It is called 'Cartesian' because it represents a program using a two-dimensional grid of nodes [Miller, 2011]. The genotype is a list of integers that represent the program primitives and how they are connected together. This section introduces CGP, describes what is CGP, how it is used, and provides suggestions of parameter settings for the CGP program.

2.2.1 General Form of CGP

The types of nodes in CGP are decided by designers and each node in the directed graph represents a particular function which is encoded by a number of genes. Figure 2.7 shows the general form of CGP.

 $F_0C_{0,0}$ $C_{0,a}F_1C_{1,0}$ $C_{1,a}$ $F_{(c+1)r_1}C_{(c+1)r_1,0}$ $C_{(c+1)r_1,a}O_0O_1$O_m

Figure 2.7: General form of CGP [Miller, 2011]

There are n_c columns and n_r rows. Furthermore, the number of program inputs is n_i and the number of outputs is n_o . Nodes in the same column are not allowed to be connected to each other. The CGP genotype of this general form is in the bottom of figure 2.7. *Fⁱ* represents function genes, $C_{i,j}$ means connection genes, and O_i in this genotype shows output genes. All functions have as many inputs as maximum function arity *a*, and unused connections are ignored.

CGP has three parameters chosen by designers which are the **number of columns**, the **number of rows** and **levels-back**. They are denoted by n_c , n_r and *l*, respectively. The product of n_c and n_r determine the maximum number of computational nodes allowed: $L_n = n_c n_r$. Levels-back, which is the parameter *l*, controls the connectivity of the graph encoded. It constrains which columns a node can get its inputs from. Varing these parameters leads to various kinds of graph topologies, e.g., tall and thin, short and wide, or layered graphs [Miller, 2011].

2.2.2 An Example of CGP Genotype-Phenotype Mapping

CGP can represent a large number of computational structure. After giving the general form, an example of a CGP genotype encoded for a digital circuit is discussed. Figure 2.8 shows the evolved CGP genotype and corresponding phenotype for a two-bit multiplier circuit. The CGP genotype parameters for this example are $n_c=10$, $n_r=1$ and $l=10$.

Figure 2.8: A CGP genotype and corresponding phenotype for a two-bit multiplier circuit [Miller, 2011]

There are four logic gates which are primitive functions in the function set. Let's assume two inputs to be *a* and *b*, and the four functions are $AND(a,b)(0)$, $AND(a,NOT(b))(1)$, $XOR(a,b)(2)$ and $OR(a,b)(3)$. The digits in parentheses are the function genes. These are the addresses in a user-defined lookup table of functions and are shown in front of each gene with an underline in the genotype. Labels 6 and 10 are not used because there are no circuit outputs requiring them, which are non-coding nodes and shown in dashed line [Miller, 2011].

2.2.3 Variation of CGP Genotype

Variation mechanisms in CGP are discussed in this subsection. Only the mutation operator is used in CGP, because the crossover operator is considered too disruptive in CGP and experience has illustrated that mutation can produce good evolutionary behaviour. How mutation works in CGP and genotype-phenotype mapping will be discussed here.

Mutation of CGP

The mutation operator used in CGP is a point mutation operator. The percentage of the total number of genes in the genotype that are mutated is defined as **mutation rate**, which is denoted by μ_r . Often the actual number of gene sites that could be mutated in a genotype is given length L_g . The quantity of mutation length is shown as μ_g , so that $\mu_g = \mu_r L_g$.

An example of the point mutation operator is shown in figure 2.9.

After mutation (b)

Figure 2.9: An example of the point mutation operator before and after it is applied to a CGP genotype [Miller, 2011]

The single point mutation applied to a CGP genotype and the corresponding phenotype occurs in the program output gene o_A changing the value from 6 to 7. This leads to nodes 3 and 7 becoming active, whereas nodes 2, 5 and 6 are changed to inactive. Dashed lines are used to represent inactive areas [Miller, 2011].

2.2.4 An Example of Cartesian Genetic Programming (CGP) in Hardware

Consider an example from [Sekanina, 2004*b*] to implement CGP in hardware. As seen in figure 2.10, the chromosome represents the configuration bits which determine the configuration of the programmable elements.

Figure 2.10: An example of CGP in Hardware [Sekanina, 2004*b*]

One CGP node consists of two multiplexers and one Configurable Functional Block (CFB). After input signals are selected by both of the 2-bit configuration bits multiplexers, they pass the signals from the multiplexers to the Configurable Functional Block (CFB). These signals are also selected by 2-bit configuration bits in the Configurable Functional Block (CFB) because four functions can be chosen in the Configurable Functional Block (CFB). Afterwards, the output signal from the Configurable Functional Block (CFB) goes to the next stage. The parameters set of this example are number of rows $n_r = 4$, number of columns $n_c = 3$, number of inputs $n_i = 4$, number of outputs $n_o = 2$, number of node inputs $n_n = 2$, number of node functions $n_f = 4$, and level-back $L = 1$.

2.2.5 Parameters Setting Suggestions of CGP

In order to make CGP appropriate for evolution, some general suggestions can be considered. Here, there are three CGP parameters, mutation rate μ_r , the number of columns n_c and the number of rows n_r , discussed below.

A suitable mutation rate depends on the length of genotype which means how many nodes there are. As a rule of thumb, one should use about 1% mutation if a maximum of 100 nodes are used, i.e. $n_c n_r = 100$ [Miller, 2011]. Generally speaking, larger genotypes require lower fitness evaluations than smaller genotypes to achieve successful evolution. The reason is the usefulness of neutral drift in the evolution of CGP genotypes. There are a lager percentage of non-coding genes in lager CGP genotypes than in smaller ones, so the potential for neutral drift is much larger. Due to this, we have seen that large genotypes lead to more efficient evolution. However, the designers need to consider the optimal number of columns *n^c* and number of rows n_r for a certain genotype length. [Miller, 2011] offers some suggestions : If there are no problems with implementing arbitrary directed graphs, then the recommended choice of these parameters is $n_r = 1$ with $n_c = l$. However, if the evolved CGP genotypes on digital devices with limited routing resources are implemented for evolving circuits, it is often useful to choose $n_c = n_r$. It should be stressed that these recommendations are rules of thumb, as no detailed quantitative work on this aspect has been published [Miller, 2011].

2.3 Evolvable Hardware (EHW)

Evolvable hardware (EHW) is a dynamic field that brings together reconfigurable hardware, artificial intelligence, fault tolerance and autonomous systems. EHW uses simulated evolution which is performed by a variety of different stochastic search algorithms such as Genetic Algorithms (GA), Evolutionary Programming (EP) or Evolution Strategies (ES) to search for new hardware configurations [Greenwood & Tyrrell, 2007*d*]. Referring to [Martin, 2003], compares the implementation of hardware using Genetic Programming (GP) with a software implementation using the same algorithm. The Hardware GP, the same algorithm running on a 1.4 GHz Athlon processor, and a standard tree based GP system using the package called lilgp running on a 1.4 GHz Athlon PC are used in the thesis. From Martin's thesis, it is obvious that the running speed of Hardware GP is much faster than the other two. There are many Evolvable Hardware (EHW) applications in the real-world such as Robot Control, Image processing, and Classification/Recognition. The target of Evolvable Hardware (EHW) will be to make better computing systems than traditional architectures for real-world applications [Torresen, 2008].

Generally speaking, there are three sorts of reconfigurable devices that evolved hardware is implemented on. These are : field programmable gate arrays (FPGAs), field programmable analog arrays (FPAAs) or field programmable transistor arrays (FPTAs). Each device depending on the function, is configured to define its architecture, and the purpose is to find the architecture of the best performance for the given application. This section provides the fundamental concepts of EHW to explain evolutionary cycle and intrinsic vs. extrinsic evolution. Furthermore the final subsection describes some information about digital and analogue reconfigurable devices.

2.3.1 Evolutionary Cycle

The concept of an evolutionary cycle is basic and important for Evolvable Hardware (EHW) and is described in this subsection. For biological evolution, lifeforms adapt to a particular environment over successive generations. The offspring come from selection, crossover, mutation and then the survival of the fittest. As for Evolutionary Computing (EC), it imitates biological evolution to optimize solutions to a wide variety of complex problems [Torresen, 2008].

Consider figure 2.11, the left side is a stylized model of the evolutionary cycle occurring for human being.

Figure 2.11: Evolutionary Cycle [Greenwood & Tyrrell, 2007*d*]

To avoid a chicken and egg argument, we create a beginning with parents to produce children. Some will survive in an environment that they are better suited to than others. Selection chooses whether these children will or will not become parents for the next generation. Moving to the right side of figure 2.11, it shows the generic electronic evolutionary cycle similar left figure but for a evolvable hardware system. Now our population is a set of electronic systems and has been transformed into a binary string which is used to configure the system. Afterwards, the system is evaluated to give a fitness to transform the problem solution into a measurable number. Using this number helps in the selection stage and in forming the next generation of the possible solution. DNA from the left figure is transformed to a binary string in the right figure. It does not have to be a binary string, but this is a good example for discussion. As the cycle illustrated in figure 2.11 and mentioned the terminal criteria in subsection 2.1.1, this cycle continues more typically until the fitness achieving a maximum value or the designers don't want to do [Greenwood & Tyrrell, 2007*d*].

2.3.2 Intrinsic vs. Extrinsic Evolution

From an implementation point of view, there are five categories of evolvable hardwarebased experiments and applications which are introduced in [Sekanina, 2004*c*]. They are named *intrinsic*, *extrinsic*, *unconstrained*, *constrained*, and *mixtrinsic* evolution. These five categories of evolution are used for evolutionary circuit design in the digital as well as the analogue domain. In this thesis only *Intrinsic* and *Extrinsic* evolution are chosen to be studied.

Figure 2.12 shows the main steps for the evolutionary synthesis of electronic circuits.

Figure 2.12: Evolutionary synthesis of electronic circuits [Greenwood & Tyrrell, 2007*d*]

There are two different ways to check fitness and which method is chosen depend on whether **intrinsic** or **extrinsic** evolution is used. Extrinsic evolution is represented as a triangle in figure 2.12 using circuit models and simulators to evaluate circuit configurations [Greenwood & Tyrrell, 2007*d*. This means that the circuits are evaluated by a software simulator. The physical reconfigurable circuit is created the configuration bitstream only from the best solution at the end of the evolution [Sekanina, 2004*c*]. On the other hand, in an intrinsic evolution, every chromosome is downloaded and physical testing to measure fitness, being a represented by a square in figure 2.12 [Greenwood & Tyrrell, 2007*d*]. Much of the circuit synthesis work done today relies on simulators because they are more flexible and it is easier to try different circuit configurations. But simulators usually do not operate very well. Many designs are chosen to be undertaken intrinsically because it leads to faster fitness calculation and faster evolution [Greenwood & Tyrrell, 2007*e*] [Sekanina, 2004*c*].

2.3.3 Reconfigurable Devices

Reconfigurable Devices are the key component of EHW. The concepts of digital and analogue reconfigurable devices are discussed in this subsection individually.

Digital Devices

For EHW, designers are interested in two types of programmable device, one being the **programmable logic device (PLD)** and the other is the **field-programmable gate** **array (FPGA)**. The field-programmable gate array (FPGA) is introduced in the following subsection.

The FPGA is a highly configurable digital device which is organized by an array of logic blocks [Greenwood & Tyrrell, 2007*f*] as shown in figure 2.13. Every FPGA chip is composed of a finite number of predefined resources containing programmable interconnects to implement a reconfigurable digit circuit. Furthermore, the I/O blocks allow the circuit to communicate with the outside world [Corporation, 2015].

Figure 2.13: Basic FPGA architecture [Corporation, 2015]

In general, there are three major types of element required on FPGAs. These are Configurable Logic Block (CLB), Interconnect, and I/O pins. First of all, the model of Configurable Logic Block (CLB) is depicted in figure 2.14.

Figure 2.14: General model of a Configurable Logic Block [Martin, 2003]

The Configurable Logic Block (CLB), also named Slice, can usually perform the function of several typical logic gates [Greenwood & Tyrrell, 2007*f*]. Each Slice includes two Logic Cells which consist of a function generator implemented as a *Look up Table* (LUT), a storage element or *Flip Flop* (FF), and an internal *Carry and Control* (CC) [Martin, 2003]. Secondly, interconnections are made between the logic elements. FPGAs typically provide several types of interconnect according to the distance between the connection of the combinational logic blocks. Finally, the I/O pins denoted as **I/O blocks (IOBs)** are generally programmable as inputs or outputs and often provide other features such as low-power or high-speed connections [Wolf, 2004]. Programming an FPGA requires three tasks: 1*⟩* the functions are implemented in logic blocks, 2 \rangle the signal is routed between logic blocks, 3 \rangle the characteristics of the input/output blocks are defined. This will be either a tri-state output or a latched input [Greenwood & Tyrrell, 2007*f*].

Analog Device

While digital hardware is becoming more and more popular, analogue electronic circuits are still required. Sensors are an example of this issue, which are usually used by analogue preprocessors to measure physical signals in an analogue world. In analogue EHW, the transistor level of the device is referred to as a **field programmable transistor array (FPTA)**, and the higher level of the device where the user programs provide operatioal amplifier circuitry is referred to as a **field programmable analogue array (FPAA)**. The abstract view of FPTA and the transistor cell are introduced in this subsection.

The picture on figure 2.15 shows a low-level structure schematic called the transistor cell where designers can vary the physical dimensions of a transistor.

Figure 2.15: Schematic of one transistor cell [Greenwood & Tyrrell, 2007*f*]

Varying the ratio of the transistor's channel Width W and Length L can produce transistors with different characteristics. A fine-grained programmable device can be constructed by placing lots of transistor cells into a $N \times M$ mesh array. The three terminals known as the Drain, the Gate and the Source of the programmable transistor can be connected to either of four cell edges, as well as to *Vdd* or Gnd. Routing signals through the chip is controlled by the six routing swithchs that link the four cell borders with each other.

The FPTA consists of 16×16 programmable transistor cells depicted as in figure 2.16.

Figure 2.16: Schematic of the FPTA [Greenwood & Tyrrell, 2007*f*]

CMOS transistors come from two types of transistors named P- and N- MOS. Half of the transistor cells are designed as programmable PMOS transistors and half as programmable NMOS transistors. Each cell contains the programmable transistor itself, three decoders that allow the three transistor terminals, Drain, Gate and Source, to be connected to one of the four cell boundaries and *Vdd* or Gnd [Greenwood & Tyrrell, 2007*f*]. As both of the analogue and digital CMOS circuits depend on the functions implemented by transistors, the FPTA is flexible enough to synthesize analogue, digital, and mixed-signal circuits [Zebulum et al., 2000].

2.4 Evolvable Hardware Architectures for Classification Tasks

There are a number of papers exhibiting classification tasks by evolvable hardware (EHW) such as pattern recognition systems applied in EHW. Facial image processing with functional unit row (FUR) is one use of the EHW architecture and is illustrated in [Glette et al., 2007]. Compared with [Yasunaga et al., 2000], which proposed Logic Design using the Evolved Truth Table (LoDETT) associated with increased Complexity Evolution (ICE) Architecture to categorize patterns, [Glette et al., 2007] exhibits slightly better classification accurancy. In addition, LoDETT lacks online evolution ability which is a drawback of the design. [Wang et al., 2005] and [Torresen et al., 2004] apply ICE architecture to implement a character classification system. This method is built on a reconfigurable structure, and the EHW system is utilized to evolve pattern classification in a short time, especially in [Torresen et al., 2004]. Electromyographic (EMG) prosthetic hand control (PHC) [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008] classifies muscles' signals, which is comparable with conventional classification techniques such as k-nearest-neighbor (kNN), decision trees (DTs) and SVMs. Also, the ECGP model and the Functional Unit Row EHW Architecture were demonstrated. SVM yields the lowest error rate, and the rate of test errors in EHW approaches is between that of kNN and DT. Packet classification with hash function [Widiger et al., 2006] classifies packets as memory and computation time are the most troublesome problems of communication networks. Furthermore, in order to limit the costs of hardware, designers used software implementations of encryption algorithms. Also, it consists of an evolutionary algorithm implemented in hardware.

[Glette, Torresen, Kaufmann & Platzner, 2008] compares EHW architectures for classification tasks. This section describes classifiers developed by conventional time sequence classification techniques such as **PLA-Based** architecture and **Increased Complexity Evolution (ICE)** architecture. Afterwards, **classification module** developed from ICE architecture consist of *P* **category detection modules (CDMs)** which are depicted in figure 2.17.

Figure 2.17: EHW classification module [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008] [Torresen et al., 2008] [Glette et al., 2007]

The input data pattern is classified by each CDM. The highest value output from the CDM will be detected by a maximum detector, and the number of this category will be the output from the system. Depending on this classification module, two classifiers are developed from it. These are the **Embedded Cartesian Genetic Programming** (ECGP) architecture and the **Functional Unit Row** (FUR) architecture. Furthermore, the measure qualities such as classification accuracy and evolution speed are required in the system. These classification architectures are described below.

2.4.1 Conventional Classifier

Three sorts of conventional classifying paradigms are selected to realize different forms of classification tasks.

• k-nearest-neighbor (kNN)

The kNN classification is one of unsupervised learning algorithms. It approaches the classification task by identifying the k-th nearest neighbours to a query example and using those neighbours to help classifying the query example which it belonged to. This method is simple, and does not require training time because of unsupervised learning. However, the problems are how to select suitable features and the setting of distances. If the designers encounter some special features, maybe the machine will measure incorrectly. On the other hand, it is difficult to set the distance unit if the case is not concerning distances. Furthermore, memory intensity is also another problem.

• Decision trees (DTs)

The DT classification is a simple supervised learning algorithm which is a diagram of nodes, terminal nodes (leaves) and decision point (branches) like a tree. Each **node** represents a test attribute, and each **branch** shows a possible test outcome. A tree is labeled either as a **leaf node** with a class or a **test node** computed from some outcome of an instance. The DT learning can be understood as a set of *if-then* rules. An input pattern is classified by successful training data from the root of a DT down to a leaf. The advantages of the DT is that the graphic approach can help decision-makers represent alternative decisions, possible outcomes, and chance events schematically. The visual approach is particularly helpful for comprehending the problems of multiple choices and possible outcomes. On the contrary, it maybe increase failures if there are too many categories.

• Support Vector Machines (SVMs)

The SVM classifier also uses a supervised learning algorithm, which builds a hyperplane or set of hyperplanes in a high or infinite dimensional space to do classification tasks. A good classification is achieved by using parallel hyperplanes with the largest distances.

These hyperplanes are optimal which means they have a maximal margin, where the margin is the minimal distance from the separating hyperplane to the nearest data points. It can deal with linearly separated problems, as well as nonlinearly ones. The key advantage of SVMs is the principle of structural risk minimization which typically yields very good generalization performance compared to other classifier paradigms [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008].

2.4.2 Taxonomy of EHW Architectures for Classification Assignment

Referring to [Glette, Torresen, Kaufmann & Platzner, 2008], four kinds of Evolvable Hardware (EHW) architectures for classification tasks are mentioned listed below in chronological order.

PLA-Based Architecture

The programmable logic array (PLA) approach implements a sum of products representation, as shown in Figure 2.18, and it is a basic model of the genome's representation.

Figure 2.18: PLA-based architecture [Glette, Torresen, Kaufmann & Platzner, 2008]

For the PLA classifier, each primary input and their negates are connected to a set of AND gates. Then, the outputs of the AND gates are connected to the inputs of a set of OR gates. A genome from the memory configures the PLA by defining the fuse setting at the intersections of primary input lines and the AND gates' inputs, the AND gates' outputs

and the OR gates' inputs. Each single OR gate's output corresponds to the detection of a single category. The fitness is measured on the correct outputs from the classifier after being presented all of the training vectors. In order to reach the maximum classification accuracy, the PLA architecture requires a high number of fitness evaluation. Furthermore, it is limited by its two-level logic structure.

Increased Complexity Evolution (ICE) Architecture

The ICE architecture can be considered as a multi-PLA that supports several sum-ofproducts for each category detection unit. The classifier system is divided into subsystems, which is seen in Figure 2.19.

Figure 2.19: Increased Complexity Evolution (ICE) Architecture [Glette, Torresen, Kaufmann & Platzner, 2008]

Each subsystem is combined as an AND-OR gate, a selector unit and a counter. The AND gates can connect to any bit of the input and the OR gates can connect to any output from the AND gates. The outputs from the OR gates are fed into the Selector which chooses which one of the outputs will be counted by the Counter. After this, all of the subsystems are connected to a Maximum Detector. Every single category is detected by a subsystem and passed to the maximum detector deciding which one of the subsystems (which category) has the highest value. The fitness is measured on the outputs of the OR layer. The concept of dividing the system into subsystems represents using smaller genomes and simpler search which may reduce the evolution time. However, it is still limited to a multiple two-level logic architecture.

The Embedded Cartesian Genetic Programming (ECGP) Architecture

Based on the **Classification Module** which is extended from ICE architecture, figure 2.20 shows the structure of the Embedded Cartesian Genetic Programming (ECGP) formed by a number of Category Detection Modules (CDMs), Summation Blocks and a Maximum Detector.

Figure 2.20: Embedded Cartesian Genetic Programming (ECGP) Architecture [Glette, Torresen, Kaufmann & Platzner, 2008]

Each CDM is a category classifier and a digital circuit evolved within the ECGP model. ECGP is an extension of the popular FPGA-oriented CGP model [Miller & Thomson, 2000]. The module size is restricted, so the maximal genome size is also limited. Compared with ICE, ECGP neither has a single or multiple two-level logic structure restriction nor a limited set of logic block functions. The drawback of this architecture is the complexity of computation.

The Functional Unit Row (FUR) Architecture

Regarded as another top-level module with ECGP, the Functional Unit Row (FUR) architecture which consists of CDMs and a Maximum Detector shown in figure 2.17. Each CDM of the FUR architecture owns $M \times N$ FUs, M N-input AND gates and one Counter illustrated on figure 2.21.

Figure 2.21: Category Detection Module [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008]

Moreover, each FUR consists of *N* FUs. The input pattern is passed to every FU, then the 1-bit outputs from the FUs in a row are fed to an *N*-input AND gate. Every output from the AND gate is connected to a counter which counts the number of activated FURs. As the number of FURs is increased, so is the output resolution from each CDM. Each FUR is evolved from an initial random bitstream, which ensures a variation in the final evolved FURs [Glette, Torresen, Kaufmann & Platzner, 2008] [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008].

Consider figure 2.22, which is the structure inside the FU being the reconfigurable element of the architecture.

Figure 2.22: Functional Unit [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008]

Each FU behaviour is controlled by a configuration line which is connected to the configuration registers. The Data Mux selects which input data can be fed into the functions f_1 and *f*2. Any number and type of functions can be adopted. Figure 2.22 illustrates an example for two functions. In addition, the unit and the input data elements are modulated by a constant value *C*. These are used by the function to compute the output from the unit. The *f* MUX is used for selecting which function leads to the output. Based on table 2.1, *I* represents the selected input value, *O* the output and *C* the constant value. The functions available to the FU input elements are chosen to be greater than or less than and equal to the constant value *C*. This means that when the input is compared with this constant value then results a true or false output. The fitness function *F* for a row can be expressed as in the following formula,

Description	Function	
Great than	$\vert O=1$ if $I > C$, else 0	
1 Less than or Equal $ O=1$ if $I \leq C$, else 0		

Table 2.1: The behaviour of the Functional Unit

where o is the output of the FUR, and each FUR is fed with all the training vectors V_t .

$$
F = \sum_{v \in V_t} x_v \qquad \text{where } x_v = \begin{cases} A \cdot o & \text{if } C_v = C_p \\ 1 - o & \text{if } C_v \neq C_p \end{cases}
$$

The fitness will be given a positive(1) output for vector v belonging to its own category $C_v = C_p$, and the value *A* is added to the fitness sum. On the contrary, negtive(0) output

will be given to the fitness for the rest of the vectors *v* which belong to the category $C_v \neq C_p$, then adding 1 to the fitness sum. Furthermore, the other cases do not contribute to the fitness value [Glette, Torresen, Kaufmann & Platzner, 2008] [Glette, Torresen, Gruber, Sick, Kaufmann & Platzner, 2008].

2.5 Summary

This chapter has introduced the area of **Evolutionary Computing (EC)**. In particular, section 2.1 renders the principles of **Evolutionary Algorithms (EAs)**. **Cartesian Genetic Programming (CGP)** is introduced in section 2.2 and it is proposed that this will be an interesting technique to be used for the research discussed in this thesis. Section 2.3 presents guidelines of **Evolvable Hardware (EHW)**. Following section 2.3, in the final section of chapter 2, four kinds of **Evolvable Hardware (EHW)** architectures for classification tasks are introduced. The next chapter considers an implementation of **Cartesian Genetic Programming (CGP)** for facial expression classification and Chapter 4 presents results of such a system.

Chapter 3

Software Implementing of CGP in Facial Expressions Classification

Following on from the background provided in the previous chapters, this chapter will consider software simulation for classifying facial expressions into sad and smiling expressions using CGP. The first section of this chapter covers how to convert input data from pictures into bit streams and how to set bit streams to buffers facilitating cross-validation. The second section explains the CGP structure used to simulate the classification results. The final section describes a Support Vector Machine (SVM) classifier implementation for facial expression classification.

3.1 Input Data Encoding

The input data for this experiment was produced by hand and consists of 24 pictures (12 smile faces and 12 sad faces) which are 8*×*8 images (64 bits). There are 20 pictures (10 smile faces and 10 sad faces) used for training data and 4 pictures (2 smile faces and 2

sad faces) used for testing data. The results will record training accuracy and test accuracy separately. The encoding directions are from top to bottom and left to right as illustrated in figure 3.1. Here an example bitstream is shown on the bottom of the figure, comprising of a total of 64 bits for one image.

0000000000000000001001000000000001000000001000010001001000001100

Figure 3.1: Input Data Encoding

Following figure 3.1, the 64-bit bitstream is divided into eight 8-bit inputs, each input representing each row of the image. In order to facilitate processing by CGP, inputs are transformed from binary to integer representations as shown in figure 3.2. For instance, the first line the top row of figure 3.1 which shows 00000000 and is transferred into the integer 0 to become the first input in[0]. In the same way, the third row is 00100100 which is transferred to the integer 36 to become the third input in[2] etc.

0000000000000000001001000000000001000000001000010001001000001100

Figure 3.2: Input Data to Integer

3.1.1 Input Data Format

The first step of the experiment uses test input data according to the order of input patterns. The order: sad1, sad2, smile1, and smile2 is selected for the first set of test data, with the rest belonging to training data. Next, the order: sad3, sad4, smile3, and smile4 becomes second set of test data. The remaining patterns form the training data. The input data are placed in buffers on the basis of a sequence so as to facilitate coding. Two examples of this k-fold cross validation illustrate the training data input of the experiment. The next section will explain what the k-fold cross validation is.

Consider figure 3.3 in which the order sad1, sad2, smile1, and smile2 is chosen to form the test data. The other 20 images from pattern 3 (sad3 and smile3) to pattern 12 (sad12 and smile12) form the training data. When $i = 0$ and j is from 3 to 11, the buffer will have 10 sad images (sad3 to sad12) in buffer[0] to buffer[9]. Then, $i = 1$ and j is also from 3 to 11, 10 smile images (smile3 to smile12) are placed in buffer[10] to buffer[19].

 $\overline{1}$

Input patterns except sad1 sad2 smile1 smile2

Figure 3.3: Input Data to Buffer example 1

The next example fetches pattern 3 (sad3 and smile3) and pattern 4 (sad4 and smile4) for test data. Figure 3.4 indicates the training set of this example. Because of discontinuous pattern inputs, two parts of the parameter *j* are used in order to be placed in the buffer sequentially. This means the pattern inputs separate into two portions. One portion contains pattern 1 (sad1 and smile1) and pattern 2 (sad2 and smile2), the other contains pattern 5 (sad5 and smile5) to pattern 12 (sad12 and smile12). Referring to figure 3.4, when $i = 0$, there are two For loops over the parameter j which are j from 1 to 2 and j from 5 to 12 so as to consecutively input sad images (sad1, sad2, sad5 to sad12) into buffer[0] to buffer[9]. Similarly when $i = 1$, the smile images smile1, smile2, and smile5 to smile12 will be placed into buffer[10] to buffer[19].

Input patterns except sad3 sad4 smile3 smile4

Figure 3.4: Input Data to Buffer example 2

There are 6 different types of input patterns in each experiment. The examples mentioned above are 2 of the 6 types, and the remaining 4 are similar. Appendix A.1 presents all the input data formats described in this section.

3.1.2 Additional Input Data Format for K-fold Cross-Validation

For the purpose of improving the accuracy of facial expression classification, additional cross validation was performed in this experiment. Selection of the testing set starts from the head and the tail of all 24 input patterns. That is pattern 1 (sad1 and smile1) and pattern 12 (sad12 and smile12) are chosen first, then pattern 2 (sad2 and smile2) and pattern 11 (sad11 and smile11). Next, pattern 3 (sad3 and smile3) and pattern 10 (sad10 and smile10) are chosen, and so on. There are 6 parts to each experiment as explained in the current and previous sections. Thus forming the method of 6-fold cross-validation. All of the cases are demonstrated in appendix A.2. Figure 3.5 shows one sample of input patterns using cross validation for training. Here example 2 is explained, the others are similar. For this example, Pattern 2 (sad2 and smile2) and pattern 11 (sad11 and smile11) are chosen as test data which are not present in the training input. There are three parts of the pattern input divided to help consecutive placing to buffers. Similarly to section 3.1.1, when $i = 0$, sad1, sad3 to sad10, and sad12 are put into buffer[0] to buffer[9]. Afterwards, smile1, smile3 to smile10, and smile12 are put into buffer[10] to buffer[19] when $i = 1$.

Input patterns except sad2 sad11 smile2 smile11

Figure 3.5: Input Data to Buffer for cross validation example 2

3.2 Algorithm Implementation

Section 3.1 considered encoding and how to convert input data from images to integer data. The simulation stages will be discussed in this section. Considering figure 3.6, the initial step involves generating random chromosomes. This is followed by reading input data. Section 3.1 goes on to describe how each input (buffer) is assigned to each tag as illustrated in figure 3.7. For 10 sad faces, every tag is assigned the value 0 while the other 10 smile faces' tags are assigned the value 1. For instance, buffer[0] is the first sad image associated with $tag[0]$ which is allocated the value 0, and buffer[10] is the first smile face input associated with tag[10] which is assigned the value 1. The process of inputting the data of the images into CGP is described in section 3.2.1 and 3.2.2. After executing the CGP, if the output value > 127 , the parameter result $= 0$ which means the classifier classifies this image as sad. Otherwise, when the output value ≤ 127 , the parameter result will be 1 implying this image should be classified as a smile. Therefore, if the values of the parameter result and tag are equal, the fitness will increase meaning that the correct classification. On the contrary, the fitness value is unchanged if the values of the result and the tag are unequal. After running the evaluation, the system will check termination conditions such as whether the EA achieves the generation limit set by the user or reaches maximum fitness. The maximum fitness is 20 corresponding to all correct classification. If the terminate conditions are accomplished, the system terminates. Otherwise, selection and mutation for the next generation operation proceeds .

Figure 3.6: The simulation flow chart

sad			smile		
buffer[0]	$tag[0] = 0$		buffer[10]	$tag[10] = 1$	
buffer[1]	$tag[1] = 0$		buffer[11]	$tag[11] = 1$	
buffer[2]	$tag[2] = 0$		buffer[12]	$tag[12] = 1$	
buffer[3]	$tag[3] = 0$		buffer[13]	$tag[13] = 1$	
buffer[4]	$tag[4] = 0$		buffer[14]	$tag[14] = 1$	
buffer[5]	$tag[5] = 0$		buffer[15]	$tag[15] = 1$	
buffer[6]	$tag[6] = 0$		buffer[16]	$tag[16] = 1$	
buffer[7]	$tag[7] = 0$		buffer[17]	$tag[17] = 1$	
buffer[8]	$tag[8] = 0$		buffer[18]	$tag[18] = 1$	
buffer[9]	$tag[9] = 0$		buffer[19]	$\text{tag}[19] = 1$	

sad

Figure 3.7: The graph of each buffer associate each tag

3.2.1 The CGP Structure for Classifying Facial Expressions

This subsection describes how CGP works in the next step of the experiment. Referring to figure 3.8, there are eight inputs to the system which are all integers in the experiments. The top of figure 3.8 is an example of a genotype for which each node consists of three numbers: the first two integers represent inputs of a node, the latter represents a function gene as shown in table 3.1. Every number below each node is the output for that node. For instance, the eighth node in figure 3.8 $\{14, 11, 5\}$ means $input(14) \times input(11)/255$. Then the result is output as the number 15. The final node of this genotype means that the overall output of the system is 26 for this example. The graph below the genotype is the corresponding phenotype possessing 1 row and 20 columns.

Figure 3.8: CGP genotype and corresponding phenotype for one example of facial expression classification

Function gene	Function definition		
$\overline{0}$	$\text{in}[0]$		
$\mathbf{1}$	$\operatorname{in}[1]$		
$\overline{2}$	$(in[0] + in[1]) % 256$		
3	$ \; \text{in}[0] - \text{in}[1] \; $		
$\overline{4}$	$(in[0] \times in[1])$ % 256		
5	$(in[0] \times in[1]) / 255$		
6	$\sqrt{in[0]+in[1]}$		
$\overline{7}$	$\sqrt{\mid in[0] - in[1] \mid}$		
8	$\mathbf{1}$		
9	255		

Table 3.1: Function set

3.2.2 Parameter Setup of CGP

In order to arrive at an optimum performance, parameters for CGP need to be considered. Table 3.2 shows the CGP parameters in these experiments. For the system described in section 3.2.1, the input consists of eight integers which produces one output. For the purpose of comparing classification accuracies, there are five different numbers of generations set which are 5,000, 10,000, 20,000, 50,000, and 100,000 generations. The population size is 5, and number of functions is 10 as indicated in table 3.1. Numbers of columns, numbers of rows, and numbers of genes per node are 20, 1, and 3 as illustrated in figure 3.8. There are 20 separate trials for the 20 training patterns in the experiment. Finally, to improve the statistical significance of all results, 20 runs of each experiment are performed.

CGP parameters	Values			
num ₋ inputs	8			
num ₋ outputs	1			
num generations	$5,000$, $10,000$, $20,000$, $50,000$, $100,000$			
population_size	5			
num_functions	10			
num _{-cols}	20			
num_rows	1			
levels_back	20			
numGenes_perNode	3			
per _c cent _c mutate	5.0			
num _{-test}	20			
num_runs_total	20			

Table 3.2: CGP parameters table

3.3 Support Vector Machine (SVM) classifier implementation by MATLAB

In order to compare the performance of a Cartesian Genetic Programming (CGP) classifier, to a more conventional classifier the Support Vector Machine (SVM) classifier is chosen. The software which is used to classify 12 smile and 12 sad images is MATLAB R2012a, and the toolbox " Bioinformatics " is needed with this MATLAB version to run the Support Vector Machine (SVM) classifier. The following paragraph will describe how to use the Support Vector Machine (SVM) classifier to classify two expression images using MATLAB.

Figure 3.9 illustrates the stages of the Support Vector Machine (SVM) classifier implementation.

Figure 3.9: The flow chart of Support Vector Machine (SVM) classifier implementation by MATLAB

First of all, 12 smile and 12 sad images are encoded to a file which forms a 24*×*64 binary data matrix as each image has 64 bits and there are 24 images in total. After encoding all facial expression images, the Cartesian Genetic Programming (CGP) classifier is used to perform two different kinds of cross validation in order to create the same comparative conditions. For example, selecting pattern1 (smile1 and sad1) and pattern2 (smile2 and sad2) to be testing patterns, leaving the others to be training patterns. Next, the MATLAB command " svmtrain " is used to generate a structure so as to do testing using the other MATLAB command " svmclassify ". Finally, the result from " svmclassify " will be outputted.

Referring to the K-fold ($K = 6$ in this thesis) cross validation in section 3.1, the second step of figure 3.9 is to select training data and test data according to section 3.1. However, the Support Vector Machine (SVM) classifier accepts binary data, so the input data does not need to be converted into an integer. In addition to selecting training data and test data, the other stages , using " svmtrain " and " svmclassify ", are all the same. The results of the experiment will be presented and explained in the chapter 4.

3.4 Summary

This chapter has described the experimental implementation. How to encode and input images to the system and do K-fold cross-validation are explained in the first section. How CGP operates to achieve an optimum performance is depicted in the latter section. In the final section, how to classify smile and sad images using the Support Vector Machine (SVM) classifier is explained. In the following chapter, the experimental results and data analysis are presented.

Chapter 4

The Experiment Results and Analysis of the Experiment

Contents

This chapter continues the simulation process of facial expression classification using Cartesian Genetic Programming (CGP) and Support Vector Machine (SVM) described in the previous chapter. The first section of this chapter will demonstrate the simulation results. The second section will analyse the performances produced by different parameters. The final section compares the performances of Cartesian Genetic Programming (CGP) classifier and Support Vector Machine (SVM) classifier.

A reminder of the set up is illustrated a schematic diagram represented in figure 4.1. There are 10 smile and 10 sad faces for training input data, and 2 smile and 2 sad faces for testing. CGP is used to assist classification of smile and sad faces. The output is one smile or one sad image which the system will check if the classification is right or wrong and calculate the fitness values.

Figure 4.1: The actual set up

The 24 images which are 8×8 pixels, 12 smile faces and 12 sad faces, which are input data described in Chapter 3 are illustrated in figures 4.2 to 4.25. Figures 4.2 to 4.13 are smile faces, and figures 4.14 to 4.25 are sad. The input data of the 12 smile faces starts here.

Figure 4.4: smile 3 Figure 4.5: smile 4

Figure 4.10: smile 9 Figure 4.11: smile 10

Figure 4.12: smile 11 Figure 4.13: smile 12

Figure 4.16: sad 3 Figure 4.17: sad 4

 $\begin{tabular}{ll} \bf Figure~4.22: \; sad \; 9 & \; \; & \; \; & \; \; \text{Figure 4.23: \; sad \; 10 \\ \end{tabular}$

Figure 4.24: sad 11 Figure 4.25: sad 12

This is the end of all the input data of the 12 smile and 12 sad faces.

4.1 Results

The results of generations against fitness are shown in Appendix B.1. The data will be analysed in next section. The experimental outcomes of cross-validation 1 described in Chapter 3.1 are demonstrated first. There are 20 runs of each parameter of generations.

In order to query the best chromosome, the graphs of the best accuracy related to each run of every generation are illustrated in Appendix B.2. The description in Appendix B.2 follows the experiments chronologically.

For the purpose of improving the readability of data, the graphs are transferred to the boxplots shown in figures 4.26 to 4.35.

Considering the example of training data over 5000 generations as shown in figure 4.26, there are 20 runs in total for each different input case. In figure 4.26, the x-axis means which pattern is excluded from the input data. For example, when x-axis shows (1,2), this means all the input data except pattern 1 (sad1 and smile1) and pattern 2 (sad2 and smile2). In other words, input data is from pattern 3 (sad3 and smile3) to pattern 12 (sad12 and smile12). Furthermore, the y-axis shows the accuracies corresponding the input patterns for 20 runs.

Figure 4.27 is relevant to figure 4.26, as it shows which means testing for 5000 generations and running 20 times as well. The x-axis represents input testing patterns such as $(1,2)$ which is input pattern 1 (sad1 and smile1) and pattern 2 (sad2 and smile2). The y-axis, the same as in figure 4.26, shows accuracies for 20 runs conforming input data.

Figure 4.26 and figure 4.27 mentioned above are two examples of the experiment. Here

indicates five different generation cases of the experiment which are set 5,000, 10,000, 20,000, 50,000, 100,000 generations respectively. Each case has two figures which show training data and testing data the same as above.

Cross validation of 5,000 generations for training 20 input patterns

Figure 4.26: 5,000 generations for 20 training input patterns

	1.2	3.4	5.6	7.8	9.10	11.12	1.12	2,11	3.10	4.9	5.8	6.7
Min	50.00%	0.00%	0.00%	0.00%	0.00%	25.00%	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Q1	75.00%	50.00%	75.00%	50.00%	0.00%	50.00%	68.75%	50.00%	50.00%	68.75%	75.00%	68.75%
Median	100.00%	87.50%	100.00%	50.00%	25.00%	50.00%	75.00%	75.00%	75,00%	75.00%	75.00%	75.00%
Q3	100.00%	100.00%	100.00%	50.00%	50.00%	50.00%	75.00%	75.00%	81.25%	75.00%	100,00%	75.00%
Max	100.00%	100.00%	100.00%	75.00%	75.00%	75.00%	100.00%	100.00%	100.00%	100.00%	100,00%	100.00%
	1.2	3,4	5.6	7.8	9.10	11.12	1.12	2.11	3.10	4.9	5.8	6.7
Min	50.00%	0.00%	0.00%	0.00%	0.00%	25.00%	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Q1-Min	25.00%	50.00%	75,00%	50,00%	0.00%	25.00%	18.75%	0.00%	0.00%	18.75%	25.00%	18.75%
Median-O1	25.00%	37.50%	25,00%	0.00%	25,00%	0.00%	6.25%	25,00%	25,00%	6.25%	0.00%	6.25%
Q3-Median	0.00%	12.50%	0.00%	0.00%	25.00%	0.00%	0.00%	0.00%	6.25%	0.00%	25.00%	0.00%
Max-O3	0.00%	0.00%	0.00%	25.00%	25.00%	25.00%	25.00%	25.00%	18.75%	25.00%	0.00%	25.00%

Cross validation of 5,000 generations for testing 4 input patterns

Figure 4.27: 5,000 generations for 4 testing input patterns

Cross validation of 10,000 generations for training 20 input patterns

Figure 4.28: 10,000 generations for 20 training input patterns

Cross validation of 10,000 generations for testing 4 input patterns

Figure 4.29: 10,000 generations for 4 testing input patterns

Cross validation of 20,000 generations for training 20 input patterns

Cross validation of 20,000 generations for testing 4 input patterns

Figure 4.31: 20,000 generations for 4 testing input patterns

1,2	3,4	5.6	7.8	9.10	11.12						
						1,12	2,11	3,10	4.9	5,8	6.7
		95.00%	95.00%	95.00%	90.00%	95.00%	95.00%	90.00%	95.00%	95.00%	95.00%
95.00%	98.75%	95.00%	98.75%	100.00%	100.00%	100.00%	98.75%	98.75%	100.00%	95.00%	100.00%
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	97.50%	100.00%
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100,00%	100.00%	100.00%	100.00%	100,00%	100.00%
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
1,2	3.4	5.6	7.8	9.10	11.12	1.12	2.11	3,10	4.9	5.8	6.7
90.00%	90.00%	95.00%	95.00%	95.00%	90.00%	95.00%	95.00%	90.00%	95.00%	95.00%	95.00%
5.00%	8.75%	0.00%	3.75%	5.00%	10.00%	5.00%	3.75%	8.75%	5.00%	0.00%	5.00%
5.00%	1.25%	5.00%	1.25%	0.00%	0.00%	0.00%	1.25%	1.25%	0.00%	2.50%	0.00%
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.50%	0.00%
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		90.00% 90.00%									

Cross validation of 50,000 generations for training 20 input patterns

Figure 4.32: 50,000 generations for 20 training input patterns

	1.2	3.4	5.6	7.8	9.10	11.12	1.12	2.11	3.10	4.9	5.8	6,7
Min	50.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	25.00%	50.00%	50.00%	50.00%	50.00%
Q1	50.00%	50.00%	50.00%	25.00%	0.00%	50.00%	75.00%	50.00%	50.00%	50.00%	75.00%	75.00%
Median	100.00%	100.00%	75,00%	50.00%	0.00%	50.00%	75.00%	75.00%	75,00%	75.00%	75.00%	75.00%
Q3	100.00%	100.00%	100.00%	50.00%	50.00%	50.00%	75.00%	75,00%	75,00%	75.00%	100.00%	75.00%
Max	100.00%	100.00%	100,00%	75,00%	75.00%	75.00%	100.00%	100,00%	100,00%	100,00%	100.00%	100.00%
	1.2	3.4	5.6	7.8	9.10	11.12	1.12	2.11	3.10	4.9	5.8	6.7
Min	50.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	25.00%	50.00%	50.00%	50.00%	50.00%
Q1-Min	0.00%	50.00%	50.00%	25,00%	0.00%	50.00%	25.00%	25,00%	0.00%	0.00%	25.00%	25.00%
Median-O1	50.00%	50.00%	25,00%	25,00%	0.00%	0.00%	0.00%	25.00%	25,00%	25,00%	0.00%	0.00%
Q3-Median	0.00%	0.00%	25.00%	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%
Max-O3	0.00%	0.00%	0.00%	25.00%	25.00%	25.00%	25.00%	25.00%	25.00%	25.00%	0.00%	25.00%

Cross validation of 50,000 generations for testing 4 input patterns

Figure 4.33: 50,000 generations for 4 testing input patterns

Cross validation of 100,000 generations for training 20 input patterns

Figure 4.34: 100,000 generations for 20 training input patterns

Figure 4.35: 100,000 generations for 4 testing input patterns

What follows are the results of the same experiment using the Support Vector Machine (SVM) described in Chapter 3. Figure 4.36 is the correct result either cross-validation 1 or cross-validation 2. For instance, the first column means cross-validation 1 for which the testing patterns are pattern 1 (smile 1 and sad 1) and pattern 2 (smile 2 and sad 2). Also, the order of the tags are smile, smile, sad and sad. In the same way, the first column of cross-validation 2 shows that the testing patterns are pattern 1 (smile 1 and sad 1) and pattern 12 (smile 12 and sad 12). The order of the tags is also smile, smile, sad and sad. The other columns of figure 4.36 have the same meanings mentioned above.

Testing Patterns						
Cross Validation 1: $(1, 2)$ $(3, 4)$ $(5, 6)$ $(7, 8)$ $(9, 10)$ $(11, 12)$						
Cross Validation 2: $(1, 12)$ $(2, 11)$ $(3, 10)$ $(4, 9)$ $(5, 8)$ $(6, 7)$						
					smile smile smile smile smile	smile
			smile smile smile smile		smile	smile
	sad	sad		sad sad	sad	sad
	sad	sad	sad	sad	sad	sad

Figure 4.36: The answer of the testing patterns of the SVM classifier

The result of cross-validation 1 is illustrated in figure 4.37, and cross-validation 2 is shown in figure 4.38.

Testing Patterns						
Cross Validation 1: $(1, 2)$ $(3, 4)$ $(5, 6)$ $(7, 8)$ $(9, 10)$ $(11, 12)$						
				smile smile smile sad	sad	sad
			smile smile smile sad		sad	sad
	sad	sad		sad smile	smile	smile
	sad	sad	sad	smile	smile	smile

Figure 4.37: The result of Cross-validation 1 of the SVM classifier

Testing Patterns

Figure 4.38: The result of Cross-validation 2 of the SVM classifier

4.2 Data Analysis

For the results of the relation between fitness and number of generations, the trend does approaches the best fitness value of the specified generation parameter as expected. However, in order to determine the best chromosome of this experiment, it is easier to use the graph and the boxplot of the best fitness value of each run. The result analysis is described as the following.

In this experiment, there are two kinds of cross-validation to be demonstrated as mentioned in Chapter 3. For the overall situation, it is obvious that the accuracies in training data are similar but cross-validation 2 is better than cross-validation 1 with the test data. Cross-validation 1 has worse accuracies because the first few input images produced are too similar with differences growing for the later images.

Comparing between 5,000 generations and 100,000 generations as shown in figures 4.26, 4.27, 4.34, and 4.35, the training data behaviour of 100,000 generations is better than of 5,000 generations. However, the test data accuracies of 5,000 generations are superior to 100,000 generations. Reason for this could be over-fitting which is training too much in 100,000 generations. This would cause the model to not learn how to generalize at all. Therefore, it is not necessarily that more generations will lead to a better performance.

Because the 10,000 generations experiment is the first attempt and 100,000 generations is the second with the same result as mentioned above, 5,000 generations, 20,000 generations,

and 50,000 generations are also executed in the experiment. Observing figure 4.26 and figure 4.28, for both cross-validation 1 or cross-validation 2 of training data, the accuracies of 10,000 generations are higher than the accuracies of 5,000 generations. However, considering the test data in figure 4.27 and figure 4.29, cross-validation 1 experiment of 10,000 generations is slightly worse than 5,000 generations. However the opposite is true in cross-validation 2. Seeing figure 4.30 and figure 4.32, the results of 50,000 generations in training data are better than the results of 20,000 generations. For the test data comparison in figure 4.31 and figure 4.33, the outcomes of 50,000 generations and 20,000 generations are evenly matched but both of them are worse than 5,000 generations and 10,000 generations. In conclusion, 10,000 generations and 5,000 generations have better accuracies in the experiment, and it is obvious that k-fold cross-validation could help for the over-fitting problem.

Another interesting issue is what is the best solution (best chromosome) of the experiment? Observing figures 4.26 to 4.35, the best solution occurs in patterns 1, 2 (sad1, smile1, sad2, smile2) of 5,000 generations. Furthermore, considering figure B.1255, the chromosome of run2 is chosen because it is more powerful to pull the accuracies from 50% to 100%. Figure 4.39 shows the genotype and phenotype of the best chromosome in the experiment.

Figure 4.39: The best chromosome of the experiment in 5,000 generations for patterns 1, 2 in run2 from figure B.1255

This circuit only has active nodes. Referring Table 3.1 and figure 4.39, the formula from the CGP description is outlined below:

$$
o_8 = |in[2] - in[6]|
$$
\n(4.1)

$$
o_9 = 255 \tag{4.2}
$$

$$
o_{10} = |in[9] - in[7]| \qquad (4.3)
$$

$$
= |o_9 - in[7]| \t\t(4.4)
$$

$$
= |255 - in[7]| \t(4.5)
$$

$$
o_{15} = in[8] \t\t(4.6)
$$

$$
= o_8 \tag{4.7}
$$

$$
= |in[2] - in[6]| \t(4.8)
$$

$$
o_{17} = (in[15] + in[10])\%256 \tag{4.9}
$$

$$
= (o_{15} + o_{10})\%256 \tag{4.10}
$$

$$
= (|in[2] - in[6]| + |255 - in[7]|)\%256 \tag{4.11}
$$

According to equation (4.11), two examples, figure 4.40 and figure 4.41 which are copies of figure 4.16 and figure 3.1, are selected randomly for validation.

Figure 4.40: sad 3 copy of figure 4.16

0000000000000000001001000000000001000000001000010001001000001100

Figure 4.41: Input Data Encoding copy of figure 3.1

The *in*[2], *in*[6] and *in*[7] of figure 4.40 are 36, 36, and 66. So the output value of figure 4.40 is (*|*36 *−* 36*|* + *|*255 *−* 66*|*)%256 = 189. This is in the range from 128 to 255 belonging to sad. The other example, figure 3.1, the *in*[2], *in*[6] and *in*[7] are 36, 18, and 12. The same way is used to calculate the output value of figure 4.41 is (*|*36 *−* 18*|* + *|*255 *−* 12*|*)%256 = 5 which locates in the range from 0 to 127 belonging to smile. The classification accuracy of the two random selection examples is 100%.

Observing equation (4.11), it shows that only *in*[2], *in*[6] and *in*[7] are considered in the best chromosome. Looking at the input images, *in*[2] is usually in the place of eyes. Furthermore, *in*[6] and *in*[7] locate in the lowest part of a mouth. The possible reason to become a best chromosome is because the locations of eyes, the lowest part of mouths have significantly different in encoding values of smile and sad images. Especially for *in*[6] and in^[7], the values of sad images are much bigger than smile images.

Depending on the experiment results mentioned before, the median accuracy rate of 5,000 generations and 10,000 generations which show in figure 4.27 and figure 4.29 are higher than 75%. Because there are only 4 test patterns used, it means only one pattern is classified as a failure. Increasing mutation rate, the numbers of testing patterns, and diversity of input patterns of one expression could improve the accuracy. CGP is a friendly algorithm to implement on hardware. By means of the software simulation, it could help to implement on hardware.

On the other hand, considering that figure 4.37 and figure 4.38 are the results of using the Support Vector Machine (SVM) classifier. According to the results of the Support Vector Machine (SVM) classifier in figure 4.36, the first three columns of cross-validation 1 experiment in figure 4.37 are correct, but the latter three columns are all wrong. Furthermore, the cross-validation 2 experiment is shown in figure 4.38. The first row and the third row are correct, but the second row and the last row are wrong. The average classification accuracy using the Support Vector Machine (SVM) classifier is 50%. Therefore, the performance of the CGP classifier is better than the Support Vector Machine (SVM) classifier in these experiments.

4.3 Summary

This chapter shows the results and explains the data of the experiments which include sad and smile expressions classification accuracies. The graphs illustrate k-fold $(k=6)$ crossvalidation of different numbers of generations involving 5,000, 10,000, 20,000, 50,000, and 100,000 generations. Section 4.2 analyses the data illustrated in section 4.1. The best chromosome is also explained there. Furthermore, this chapter also demonstrates the results and data analysis using the Support Vector Machine (SVM) classifier to show that the performance using the CGP classifier is better than using the Support Vector Machine (SVM) classifier. The next chapter will propose how to implement the idea on Evolvable Hardware (EHW).

Chapter 5

Proposed Implementation of Evolvable Hardware (EHW)

Contents

This chapter considers how the structures and techniques discussed in previous chapters might be implemented in a hardware system. The Xilinx University Program (XUP) Board is used for this experiment which has one FPGA. Other peripherals are described below. The next section explains the heuristic algorithm for EHW on FPGA. The Hardware Description Language VHDL is also studied to develop the EvoBlock which is used for customer design. Finally, an implementation of facial expression classification by CGP on FPGA will be proposed. This includes the flow chart and the Message Sequence Chart (MSC) of the proposed hardware implementation by CGP including input cell arrays.

5.1 Introduction of Xilinx University Program (XUP) Board

The Xilinx University Program (XUP) manufactured by Digilent is the demo board adopted from the beginning of this work. This contains a Virtex-2 Pro FPGA with a MicroBlaze core, an embedded platform cable USB configuration port, an on-board 10/100 ethernet PHY device and a RS-232 DB9 serial port. See [*XUPV2P_UserGuide_v1.2*, n.d.] for more on the XUP board. Configuration data is written to the FPGA via the USB port, and a USB cable is used to connect the board to the extension socket above the PC monitor. Data from the FPGA will be delivered to the PC by the RS232 serial port. In order to test communication with the board, it is required to configure the FPGA with a simple test circuit, which is programmed to run a test bit file to flash LEDs. iMPACT is the Xilinx program used to configure FPGAs. If the configuration passes, the program will output "Program Succeeded" on the iMPACT. The evolutionary system has been developed by adopting Xilinx's EDK software which allows designers to create embedded systems with both hardware and software. Next, the EDK was used on the XPS platform and an EDK block diagram of the hardware evolution system on the FPGA was generated.

The FPGA on the XUP board was created using a MicroBlaze soft core to execute the Evolutionary Algorithm (EA). The EvoBlock is then used for evaluating candidate solutions. Then a serial port called an Universal Asynchronous Receiver/Transmitter (UART) enabled the Evolutionary Algorithm (EA) to send the results back to the PC as shown as figure 5.1.

Figure 5.1: The embedded hardware evolution system [Trefzer, 2011]

The MicroBlaze microprocessor, which can be seen at the top of the diagram, and a number of blocks are connected to the core by a set of buses such as a Data Local Memory Bus (DLMB), an Instruction Local Memory Bus (ILMB), and a Memory Bus-Processor Local Bus (MB-PLB) which link the memory and the processor. In addition, the EvoBlock named *plb-evoblock* is used by customer to design the evolutionary system.

Similar to the XUP Virtex-2 Pro Development System board, an upgraded **Xilinx XUPV5-LX110T** board is used for this discussion. This board is a general purpose development board powered by the Virtex-5 FPGA. It is also a feature-rich general purpose evaluation and development platform, and includes on-board memory and industry standard connectivity interfaces, and delivers a versatile development platform for embedded applications [*Xilinx XUPV5-LX110T Evalution Platform*, n.d.].

5.2 Heuristic Algorithm of EHW on FPGA

The flow chart of the evolvable algorithm is shown in Figure 5.2, shows the heuristic sequence of EHW from the initialisation step which sets up beginning environment of system.

Figure 5.2: Heuristic algorithm of EHW on FPGA

Evaluation is the next step where the fitness function is evaluated and signals are sent to the

connected EvoBlock. This is following by selection and variation, then the system goes back to evaluation to form a cycle. The four procedures are executed on the MicroBlaze processor. The EvoBlock is configured by processor and the results are outputted from MicroBlaze.

5.3 Proposed Implementation of Facial Expression Classification by CGP on FPGA

This section proposes the idea about how to use CGP to implement facial expression classification on FPGA. Firstly, the flow chart and Message Sequence Chart (MSC) of proposed hardware implementation is illustrated. Individual parts, such as input call arrays and CGP structure hardware implementation, are explained in the following subsections.

5.3.1 The Flow Chart and Message Sequence Chart (MSC) of Proposed Hardware Implementation of Facial Expression Classification by CGP

Referring back to software simulation mentioned in Chapter 3, figure 5.3 proposes the idea of hardware implementation.

Figure 5.3: The flow chart of hardware architecture

Each input image is given an associated FPGA tag. Smile faces' values are rated from 0 to 127 and have tag values of 1, and sad faces' values are rated from 128 to 255 and have tag values of 0. Passing through the cell array described in section 5.3.2 and CGP and tags described in section 5.3.3, the result will generate a 1-bit answer. This answer means that the input images are classified by CGP and are either smile or sad expressions. Then, an exclusive NOR is used on the tags associated with each input image to return a value. If the values are the same, that means the classification is right and the output of the exclusive NOR will be 1. Otherwise, the output of the exclusive NOR will be 0 which means the classification is wrong and the fitness value will be unchanged.

The next stage after the hardware structure is designing, the data flow which is represented by a Message Sequence Chart (MSC) as shown in figure 5.4.

Figure 5.4: The Message Sequence Chart (MSC) of hardware architecture

The x-axis of the Massage Sequence Chart (MSC) shows the process from figure 5.3 including MicroBlaze, input data, tag₋ original, cell array, Cartesian Genetic Programming (CGP), tag CGP, and fitness calculation. The y-axis indicates the time sequence where the lower the position is, the later the procedure is executed. Firstly, the tag is transmitted to tag original stored in RAM, then the image data associated with the same tag is sent to the cell array in FPGA. When the write-enable signal of the cell array is triggered by a clock, the input data is delivered to Cartesian Genetic Programming (CGP). After the MicroBlaze dispatches the configuration bits to do the Cartesian Genetic Programming (CGP) procedure of classification, a one bit answer will be sent to tag CGP in RAM. Afterwards, tag CGP and tag original are transmitted to an exclusive NOR gate which will evaluate whether the answer from Cartesian Genetic Programming (CGP) is the same as the original tag or not. If these two tags are the same, the output of the exclusive NOR will be 1. If not, the output of the exclusive NOR will be 0. Finally, the answer from the exclusive NOR is delivered back to MicroBlaze. The procedure runs until the terminating condition satisfied.

5.3.2 Input Cell Arrays

When classifying facial expression images, image data input of hardware should be considered in advance. Figure 5.5 illustrates how to input image data from computer to FPGA.

Figure 5.5: The input cell arrays of hardware architecture

Before entering the Evoblock, use_{-logic}.vhdl is constructed because MicroBlaze only has 32-bit buses. Two vector registers are created with each vector register storing 32-bit of data. Considering section 3.1 on input data encoding, each image is encoded to 8 vectors and each vector has 8 bits. Since MicroBlaze is a Big endian processor, the bit sequence is sorted in

reverse on the data bus. VectorReg1 stores vector 0 to vector 3 in bit 32 to bit 63. On the other hand, VectorReg0 keeps vector 4 to vector 7 which constitute bit 0 to bit 31.

After arranging the input data in use logic.vhdl, in EvoBlock, the data in VectorReg1 is recognized as vectors V0 to V3 and will be sent to Cell Array0 to Cell Array3. In the same way, VectorReg0 delivers V4 to V7 to Cell Array4 to Cell Array7. Then, the binary vectors are transferred to integers and saved in output registers.

5.3.3 Proposed Hardware Implementation using CGP

This section describes a CGP implementation on hardware layer by layer. Firstly, the top view of CGP structure is illustrated. Next the structure of one CGP node is explained. Finally, the tag. CGP structure is described.

The Structure of CGP

The CGP structure shown in figure 5.6 has 20 columns and 1 row. So there are 20 nodes in one chromosome in total.

Figure 5.6: The CGP of hardware architecture

The configuration address (conf. addr) means which node of CGP is active. Because there is a total of 20 nodes, the configuration address is set to be 5 bits in order to enable all nodes.

The configuration data (conf. data) is given 10 bits for each node which will be explained in the next section.

The Structure of one CGP Node

The picture of figure 5.7 shows the hardware architecture of one node of CGP.

Figure 5.7: One node of the CGP of hardware architecture

This is the configuration data (conf. data) from the last section of figure 5.6. After enabling one node of CGP, the configuration data (conf. data) will be inputted. This consists of 2 data multiplexers (data MUXs). Next, both of the data multiplexers (data MUXs) connect to one function multiplexer (f MUX). Each node has 2 inputs which come from data multiplexers (data MUXs) each of which have 8 inputs. In this way, each data multiplexer (data _ MUX) possesses 3 bits selects. Furthermore, there are 10 function sets demonstrated in table 3.1, so 4 bits selects in function multiplexer (f MUX) are given. Consequently, there are 10 bits in configuration data (conf₋ data) for one node of CGP.

The Strcture of a Tag

The hardware structure of a tag₋ CGP coming from Cartesian Genetic Programming (CGP) calculation shows in figure 5.8.

Figure 5.8: The tag- CGP of hardware architecture

For the design purpose, a tag only has 2 values. When the input is from 0 to 127, the output of the tag is 1. On the contrary, the output of a tag is 0 when the input is from 128 to 255.

5.4 Summary

This chapter proposes the idea of hardware implementation referring to the software simulation in Chapter 3. The Xilinx University Program (XUP) Board, the heuristic evolutionary algorithm of EHW on FPGA, and the VHDL approach are used to practice how to implement the experiment in advance. The final section goes on to use CGP classifying facial expressions on FPGA which involves the flow chart, Message Sequence Chart (MSC), the input cell arrays, and the hardware of Cartesian Genetic Programming (CGP).

Chapter 6

Conclusions and Future Work

Contents

This chapter is the final stage and contains conclusions and future work. The first section describes the overall contribution of this thesis. It concludes all chapters mentioned previously in the thesis. The last section describes future work of using CGP to classify emotional responses and lie detection which could be implemented effectively on an Evolvable Hardware (EHW) platform.

6.1 Summary of The Contribution

There are some applications of Cartesian Genetic Programming (CGP) like image filters or CGP accelerator implemented in the FPGA [Miller, 2011]. Using CGP for facial expression classification is an innovative application. The preliminary experiment in this thesis which uses CGP to classify two different facial expressions of 64 bits data input has good results. This could be further implemented in the Field-Programmable Gate Array (FPGA).

Chapter 5 of this thesis proposes the process and architectures of classifying the facial expressions using CGP implemented in a FPGA. The final stage of the project is to make an embedded system device of a facial expression polygraph which may be of use for social work applications. The device may also detect subjects from the video which may protect social workers from threatening behaviour, for example. Therefore, the idea of the facial expression polygraph could make an important contribute to social work.

6.2 Conclusion

The project plan, as the motivation for the first chapter, is to design a device which helps social workers to distinguish whether subjects are telling truth or not from their expressions. The method was to use **Evolutionary Computing**, which is expected to perform effectively when differentiating facial expressions. Depending on the target, the following chapters explain this progressively.

The second chapter is firstly a literature review introducing the **Evolutionary Algorithm (EA)**. **Cartesian Genetic Programming (CGP)** is the method used in the experiment. The experiment is expected to be implemented on **Evolvable Hardware (EHW)**. Furthermore, doing the classification task on **Evolvable Hardware (EHW)** is the purpose of the experiment. Therefore, the background of **Evolvable Hardware (EHW)** and the literature review of **Evolvable Hardware (EHW)** architectures for classification tasks make up the last two sections of chapter 2.

Chapter 3 describes software simulation to classify 64-bit smile and sad images. There are 10 smile images and 10 sad images for training and 4 smile images and 4 sad images for testing. Input data encoding and K-fold (K=6) cross-validation for the purpose of improving accuracy are presented in section 3.1. Section 3.2 shows the use of smile and sad images classification using CGP, and parameter setup of CGP in the experiment.

The results and data analysis are demonstrated in chapter 4. Section 4.1 exhibits the whole story in the experiment from 10,000 generations extending to 100,000 generations, including 5,000 generations, 20,000 generations, and 50,000 generations. Based on the data produced from section 4.1, generally speaking, 10,000 generations and 5,000 generations have better accuracies of classification. On the other hand, section 4.2 also discusses an interesting topic of the best solution (best chromosome) in the experiment which occurs when using patterns 1, 2 (sad1, smile1, sad2, smile2) over 5,000 generations. The solution and validation are shown in the section 4.2.

Chapter 5 proposes ideas of hardware architectures without implementation. The first three sections introduce the Xilinx University Program (XUP) Board, and heuristic algorithm of EHW on FPGA, as preparations for hardware implementation. Section 5.3 considers feasible proposals including the flow chart of hardware implementation, input cell arrays, and CGP architectures.

6.3 Future Work

The final form of input in the project is video. In this procedure, multi-chromosome CGP may be used due to the complicated and large amount of data from the video needed to be handled. The video is about emotional responses and lie detection through facial expression features. A multi-chromosome genotype is divided up into *n* equal length chromosomes. The number of chromosomes is dictated by the number of outputs of a given problem, and each chromosome has a single output. The entire problem is still represented in a single genotype. A recommendation of this case could be to adopt each feature to become each singlechromosome from one picture. Therefore, one picture may consist of a multi-chromosome for all features. Assuming the data type of the features is recognised by the system, then processing can be performed using a parallel method by multi-chromosome CGP.

According to the function of multi-chromosome CGP, a novel Evolvable Hardware (EHW) structure required to be developed. It probably has a large amount of data and higher resolution requirement, but CGP could be implemented to carry out the classification task on EHW. VHDL is the language which will be used to develop the hardware structure.

This study is expected to evolve the back-end of a polygraph, which is a classifier for facial expressions. The users will be social workers in order to identify suspected abusers. The first objective of this study is high accuracy, then improving the classification speed. It is hoped to make some contribution to social work.

Appendix A

Input Data Format

Contents

As described input data format in section 3.1, appendix A.1 will present all the input format of the experiment at the beginning. Following, appendix A.2 shows input data format for cross validation discussed in section 3.1.2.

A.1 Input Data Format for 3.1.1

Here represents all input data formats of training sets which are indicated 2 sorts of 6 in section 3.1.1 show in figure A.1 to figure A.6.

A.2 Input Data Format for 3.1.2

The cross validation of all input data formats which describes one example in section 3.1.2 show in figure A.7 to figure A.12.

Input patterns except sad1 sad2 smile1 smile2

Figure A.1: Input Data to Buffer example 1

pattern	buffer		
$\mathbf{1}$	(sad) Ω	(smile) 10	
2	$\mathbf{1}$	11	$i = 0$ $\frac{1}{3}$ sad $\frac{1}{\sinh$ smile $i = 1$
5	2	12	
6	3	13	for ($i = 0 - 1$)
7	$\overline{4}$	14	for $(i = 1 - 2)$ buffer $[i * 10 + (j - 1)]$
8	5	15	for $(j = 5 \sim 12)$
9	6	16	buffer $[i * 10 + (j - 3)]$
10	7	17	
11	8	18	
12	9	19	

Figure A.2: Input Data to Buffer example 2

Input patterns except sad5 sad6 smile5 smile6

Figure A.3: Input Data to Buffer example 3

pattern	buffer		
$\mathbf{1}$	(sad) Ω	(smile) 10	
2	$\mathbf{1}$	11	$i = 0$ $\frac{1}{3}$ sad $i = 1$ $\frac{\pi}{2}$ smile
3	2	12	
$\overline{4}$	3	13	for ($i = 0 - 1$)
5	4	14	for $(i = 1 - 6)$ buffer $[i * 10 + (j - 1)]$
6	5	15	for $(i = 9 \sim 12)$
9	6	16	buffer $[i * 10 + (j - 3)]$
10	7	17	
11	8	18	
12	9	19	

Figure A.4: Input Data to Buffer example 4

Input patterns except sad9 sad10 smile9 smile10

Figure A.5: Input Data to Buffer example 5

pattern	buffer		
$\mathbf{1}$	(sad) $\mathbf{0}$	(smile) 10	
$\overline{2}$	$\mathbf{1}$	11	$i = 0$ $\frac{1}{3}$ sad $i = 1$ $\frac{1}{3}$ smile
3	2	12	
$\overline{4}$	3	13	for $(i = 0 \sim 1)$
5	4	14	for $(j = 1 \sim 10)$ buffer $[i * 10 + (j - 1)]$
6	5	15	
7	6	16	
8	7	17	
9	8	18	
10	9	19	

Figure A.6: Input Data to Buffer example 6

Input patterns except sad1 sad12 smile1 smile12

Figure A.7: Input Data to Buffer for cross validation example 1

pattern	buffer		
$\mathbf{1}$	(sad) $\mathbf{0}$	(smile) 10	
3	$\mathbf{1}$	11	$i = 0$ $\frac{1}{3}$ sad $i = 1$ $\frac{1}{\sinh$ smile
$\overline{4}$	2	12	
5	3	13	for $(i = 0 \sim 1)$
6	$\overline{4}$	14	$i = 1$ buffer $[i * 10 + 0]$
7	5	15	for $(i = 3 \sim 10)$
8	6	16	buffer $[i * 10 + (j - 2)]$
9	7	17	$i = 12$ buffer $\lceil i * 10 + 9 \rceil$
10	8	18	
12	9	19	

Figure A.8: Input Data to Buffer for cross validation example 2

Input patterns except sad3 sad10 smile3 smile10

Figure A.9: Input Data to Buffer for cross validation example 3

pattern	buffer		
	(sad)	(smile)	
1	0	10	
2	$\mathbf{1}$	11	$\frac{1}{3}$ sad $i = 0$ $\frac{\pi}{2}$ smile $i = 1$
3	2	12	
5	3	13	for $(i = 0 \sim 1)$
6	4	14	for $(i = 1 - 3)$ buffer $[i * 10 + (j - 1)]$
7	5	15	for $(i = 5 \sim 8)$
8	6	16	buffer $[i * 10 + (j - 2)]$
10	7	17	for ($j = 10 \sim 12$) buffer $[i * 10 + (j - 3)]$
11	8	18	
12	9	19	

Figure A.10: Input Data to Buffer for cross validation example 4

Input patterns except sad5 sad8 smile5 smile8

Figure A.11: Input Data to Buffer for cross validation example 5

pattern	buffer		
	(sad) $\mathbf 0$	(smile) 10	
$\mathbf{1}$			
2	$\mathbf{1}$	11	$\frac{1}{3}$ sad $i = 0$ $\frac{1}{\sinh$ smile $i = 1$
3	2	12	
4	3	13	for ($i = 0 - 1$)
5	4	14	for $(i = 1 - 5)$ buffer $[i * 10 + (j - 1)]$
8	5	15	for $(i = 8 \sim 12)$
9	6	16	buffer $[i * 10 + (j - 3)]$
10	7	17	
11	8	18	
12	9	19	

Figure A.12: Input Data to Buffer for cross validation example 6

Appendix B

The Results of Experiments

Contents

Here the results described in section 4.1 are shown. Appendix B.1 shows the results of generations against fitness. Appendix B.2 presents the best accuracy (fitness) related to each run of every generation.

B.1 The results of generations against fitness

The experimental outcomes of generations against fitness are represented here. There are 20 runs of each parameter of generations. Figures B.1 to B.20 are runs 0 to 19 of training showing the fitness over 5000 generations using testing patterns 1 (smile 1 and sad 1) and 2 (smile 2 and sad 2).

Figure B.21 to B.40 are runs 0 to 19 of training showing the fitness over 5000 generations using the testing patterns 3 (smile 3 and sad 3) and 4 (smile 4 and sad 4).

Figures B.41 to B.60 are runs 0 to 19 of training showing the fitness over 5000 generations using the testing patterns 5 (smile 5 and sad 5) and 6 (smile 6 and sad 6).

Figures B.61 to B.80 are runs 0 to 19 of training showing the fitness over 5000 generations using testing patterns 7 (smile 7 and sad 7) and 8 (smile 8 and sad 8).

Figures B.81 to B.100 are runs 0 to 19 of training showing the fitness over 5000 generations using testing patterns 9 (smile 9 and sad 9) and 10 (smile 10 and sad 10).

Figure B.101 to B.120 are runs 0 to 19 of training showing the fitness over 5000 generations using testing patterns 11 (smile 11 and sad 11) and 12 (smile 12 and sad 12).

Here starts the 10,000 generations experiment for cross-validation 1. Firstly, figures B.121 to B.140 shows runs 0 to 19 of fitness using testing patterns 1 (smile 1 and sad 1) and 2 (smile 2 and sad 2).

Figures B.141 to B.160 are runs 0 to 19 of training showing the fitness over 10,000 generations using testing patterns 3 (smile 3 and sad 3) and 4 (smile 4 and sad 4).

Figures B.161 to B.180 are runs 0 to 19 of training showing the fitness over 10,000 generations using testing patterns 5 (smile 5 and sad 5) and 6 (smile 6 and sad 6).

Figures B.181 to B.200 are runs 0 to 19 of training showing the fitness over 10,000 generations using testing patterns are 7 (smile 7 and sad 7) and 8 (smile 8 and sad 8).

Figures B.201 to B.220 are runs 0 to 19 of training showing the fitness over 10,000 generations using testing patterns 9 (smile 9 and sad 9) and 10 (smile 10 and sad 10).

Figures B.221 to B.240 are runs 0 to 19 of training showing the fitness over 10,000 generations using testing patterns 11 (smile 11 and sad 11) and 12 (smile 12 and sad 12).

Here starts cross-validation 1 of the 20,000 generations experiment. Figures B.241 to B.260 are runs 0 to 19 of training showing the fitness over 20,000 generations using testing patterns 1 (smile 1 and sad 1) and 2 (smile 2 and sad 2).

Figures B.261 to B.280 are runs 0 to 19 of training showing the fitness over 20,000 generations using testing patterns 3 (smile 3 and sad 3) and 4 (smile 4 and sad 4).

Figures B.281 to B.300 are runs 0 to 19 of training showing the fitness over 20,000 generations using testing patterns 5 (smile 5 and sad 5) and 6 (smile 6 and sad 6).

Figures B.301 to B.320 are runs 0 to 19 of training showing the fitness over 20,000 generations using testing patterns 7 (smile 7 and sad 7) and 8 (smile 8 and sad 8).

Figures B.321 to B.340 are runs 0 to 19 of training showing the fitness over 20,000 generations using testing patterns 9 (smile 9 and sad 9) and 10 (smile 10 and sad 10).

Figures B.341 to B.360 are runs 0 to 19 of training showing the fitness over 20,000 generations using testing patterns 11 (smile 11 and sad 11) and 12 (smile 12 and sad 12).

Cross-validation 1 of the 50,000 generations experiment starts here. Figures B.361 to B.380 are runs 0 to 19 of training showing fitness over 50,000 generations using testing patterns 1 (smile 1 and sad 1) and 2 (smile 2 and sad 2).

Figures B.381 to B.400 are runs 0 to 19 of training showing fitness over 50,000 generations using testing patterns 3 (smile 3 and sad 3) and 4 (smile 4 and sad 4).

Figures B.401 to B.420 are runs 0 to 19 of training showing fitness over 50,000 generations using testing patterns 5 (smile 5 and sad 5) and 6 (smile 6 and sad 6).

Figures B.421 to B.440 are runs 0 to 19 of training showing fitness over 50,000 generations using testing patterns 7 (smile 7 and sad 7) and 8 (smile 8 and sad 8).

Figures B.441 to B.460 are runs 0 to 19 of training showing fitness over 50,000 generations using testing patterns 9 (smile 9 and sad 9) and 10 (smile 10 and sad 10).

Figures B.461 to B.480 are runs 0 to 19 of training showing fitness over 50,000 generations using testing patterns 11 (smile 11 and sad 11) and 12 (smile 12 and sad 12).

Here starts the cross-validation 1 experiment of 100,000 generations. Figures B.481 to B.500 are training runs showing fitness over 100,000 generations using testing patterns 1 (smile 1 and sad 1) and 2 (smile 2 and sad 2).

Figures B.501 to B.520 are training runs showing fitness over 100,000 generations using testing patterns 3 (smile 3 and sad 3) and 4 (smile 4 and sad 4).

Figures B.521 to B.540 are training runs showing fitness over 100,000 generations using testing patterns 5 (smile 5 and sad 5) and 6 (smile 6 and sad 6).

Figures B.541 to B.560 are training runs showing fitness over 100,000 generations using testing patterns 7 (smile 7 and sad 7) and 8 (smile 8 and sad 8).

Figures B.561 to B.580 are training runs showing fitness over 100,000 generations fitness using testing patterns 9 (smile 9 and sad 9) and 10 (smile 10 and sad 10).

Figures B.581 to B.600 are training runs showing fitness over 100,000 generations using testing patterns 11 (smile 11 and sad 11) and 12 (smile 12 and sad 12).

It starts the results of fitness for the cross-validation 2 experiment follows. Figures B.601 to B.620 are training runs showing fitness over 5,000 generations using testing patterns 1 (smile 1 and sad 1) and 12 (smile 12 and sad 12).

Figures B.621 to B.640 are training runs showing fitness over 5,000 generations using testing patterns 2 (smile 2 and sad 2) and 11 (smile 11 and sad 11).

Figures B.641 to B.660 are training runs showing fitness over 5,000 generations using testing patterns 3 (smile 3 and sad 3) and 10 (smile 10 and sad 10).

Figures B.661 to B.680 are training runs showing fitness over 5,000 generations using testing patterns 4 (smile 4 and sad 4) and 9 (smile 9 and sad 9).

Figures B.681 to B.700 are training runs showing fitness over 5,000 generations using testing patterns 5 (smile 5 and sad 5) and 8 (smile 8 and sad 8).

Figures B.701 to B.720 are training runs showing fitness over 5,000 generations using testing patterns 6 (smile 6 and sad 6) and 7 (smile 7 and sad 7).

Figures B.721 to B.740 are training runs showing fitness over 10,000 generations using testing patterns 1 (smile 1 and sad 1) and 12 (smile 12 and sad 12).

Figures B.741 to B.760 are training runs showing fitness over 10,000 generations using testing patterns 2 (smile 2 and sad 2) and 11 (smile 11 and sad 11).

Figures B.761 to B.780 are training runs showing fitness over 10,000 generations using testing patterns 3 (smile 3 and sad 3) and 10 (smile 10 and sad 10).

Figures B.781 to B.800 are training runs showing fitness over 10,000 generations using testing patterns 4 (smile 4 and sad 4) and 9 (smile 9 and sad 9).

Figures B.801 to B.820 are training runs showing fitness over 10,000 generations using testing patterns 5 (smile 5 and sad 5) and 8 (smile 8 and sad 8).

Figures B.821 to B.840 are training runs showing fitness over 10,000 generations using testing patterns 6 (smile 6 and sad 6) and 7 (smile 7 and sad 7).

Figures B.841 to B.860 are training runs showing fitness over 20,000 generations using testing patterns 1 (smile 1 and sad 1) and 12 (smile 12 and sad 12).

Figures B.861 to B.880 are training runs showing fitness over 20,000 generations using testing patterns 2 (smile 2 and sad 2) and 11 (smile 11 and sad 11).

Figures B.881 to B.900 are training runs showing fitness over 20,000 generations using testing patterns 3 (smile 3 and sad 3) and 10 (smile 10 and sad 10).

Figures B.901 to B.920 are training runs showing fitness over 20,000 generations using testing patterns 4 (smile 4 and sad 4) and 9 (smile 9 and sad 9).

Figures B.921 to B.940 are training runs showing fitness over 20,000 generations using testing patterns 5 (smile 5 and sad 5) and 8 (smile 8 and sad 8).

Figures B.941 to B.960 are training runs showing fitness over 20,000 generations using testing patterns 6 (smile 6 and sad 6) and 7 (smile 7 and sad 7).

Figures B.961 to B.980 are training runs showing fitness over 50,000 generations using testing patterns 1 (smile 1 and sad 1) and 12 (smile 12 and sad 12).

Figures B.981 to B.1000 are training runs showing fitness over 50,000 generations using testing patterns 2 (smile 2 and sad 2) and 11 (smile 11 and sad 11).

Figures B.1001 to B.1020 are training runs showing fitness over 50,000 generations using testing patterns 3 (smile 3 and sad 3) and 10 (smile 10 and sad 10).

Figures B.1021 to B.1040 are training runs showing fitness over 50,000 generations using testing patterns 4 (smile 4 and sad 4) and 9 (smile 9 and sad 9).

Figures B.1041 to B.1060 are training runs showing fitness over 50,000 generations using testing patterns 5 (smile 5 and sad 5) and 8 (smile 8 and sad 8).

Figures B.1061 to B.1080 are training runs showing fitness over 50,000 generations using testing patterns 6 (smile 6 and sad 6) and 7 (smile 7 and sad 7).

Figures B.1081 to B.1100 are runs 0 to 19 of training showing fitness over 100,000 generations using the testing patterns 1 (smile 1 and sad 1) and 12 (smile 12 and sad 12).

Figures B.1101 to B.1120 are runs 0 to 19 of training showing fitness over 100,000 generations using the testing patterns 2 (smile 2 and sad 2) and 11 (smile 11 and sad 11).

Figures B.1121 to B.1140 are runs 0 to 19 of training showing fitness over 100,000 generations using the testing patterns 3 (smile 3 and sad 3) and 10 (smile 10 and sad 10).

Figures B.1141 to B.1160 are runs 0 to 19 of training showing fitness over 100,000 generations using the testing patterns 4 (smile 4 and sad 4) and 9 (smile 9 and sad 9).

Figures B.1161 to B.1180 are runs 0 to 19 of training showing fitness over 100,000 generations using the testing patterns 5 (smile 5 and sad 5) and 8 (smile 8 and sad 8).

Figures B.1181 to B.1200 are runs 0 to 19 of training showing fitness over 100,000 generations using the testing patterns 6 (smile 6 and sad 6) and 7 (smile 7 and sad 7).

B.2 The Best Accuracy Related to Each Run of Every Generation

In order to query the best chromosome, the graphs of the best accuracy (fitness) related to each run of every generation are illustrated. The description here follows the experiments chronologically.

To start with, the number of generations was set to 10,000 and pattern 11 (sad11 and smile11) and pattern 12 (sad12 and smile12) were chosen as testing data. As the accuracy

was not good enough, cross-validation was adopted, and the results are shown in the graphs for training data and testing data displayed in figures B.1201 to B.1212. The x-axis shows which run of the 10,000 generations experiment and has a total range of 20 runs. The y-axis shows the accuracy for the best generation of each run.

Figure B.1201: Cross-validation 1 for 10,000 generations training data except patterns 1, 2

Figure B.1202: Cross-validation 1 for 10,000 generations training data except patterns 3, 4

Figure B.1203: Cross-validation 1 for 10,000 generations training data except patterns 5, 6

Figure B.1204: Cross-validation 1 for 10,000 generations training data except patterns 7, 8

Figure B.1205: Cross-validation 1 for 10,000 generations training data except patterns 9, 10

Figure B.1206: Cross-validation 1 for 10,000 generations training data except patterns 11, 12

Figure B.1207: Cross-validation 1 for 10,000 generations testing data patterns 1, 2

Figure B.1208: Cross-validation 1 for 10,000 generations testing data patterns 3, 4

Figure B.1209: Cross-validation 1 for 10,000 generations testing data patterns 5, 6

Figure B.1210: Cross-validation 1 for 10,000 generations testing data patterns 7, 8

Figure B.1211: Cross-validation 1 for 10,000 generations testing data patterns 9, 10

Figure B.1212: Cross-validation 1 for 10,000 generations testing data patterns 11, 12

Analysing accuracies of the 10,000 generations testing data as shown in figures B.1207 to B.1212, most results are lower than 50% which is not acceptable. Therefore, K-fold (K = 6 in this experiment) cross-validation 2 of 10,000 generations as explained in chapter 3 was used the results of which are illustrated in figures B.1213 to B.1224.

Figure B.1213: Cross-validation 2 for 10,000 generations training data except patterns 1, 12

Figure B.1214: Cross-validation 2 for 10,000 generations training data except patterns 2, 11

Figure B.1215: Cross-validation 2 for 10,000 generations training data except patterns 3, 10

Figure B.1216: Cross-validation 2 for 10,000 generations training data except patterns 4, 9

Figure B.1217: Cross-validation 2 for 10,000 generations training data except patterns 5, 8

Figure B.1218: Cross-validation 2 for 10,000 generations training data except patterns 6, 7

Figure B.1219: Cross-validation 2 for 10,000 generations testing data patterns 1, 12

Figure B.1220: Cross-validation 2 for 10,000 generations testing data patterns 2, 11

Figure B.1221: Cross-validation 2 for 10,000 generations testing data patterns 3, 10

Figure B.1222: Cross-validation 2 for 10,000 generations testing data patterns 4, 9

Figure B.1223: Cross-validation 2 for 10,000 generations testing data patterns 5, 8

Figure B.1224: Cross-validation 2 for 10,000 generations testing data patterns 6, 7

Comparing with testing data from figures B.1207 to B.1212, K-fold $(K=6)$ cross-validation obviously provides higher accuracies illustrated in figures B.1219 to B.1224. However, the results achieving 100% accuracy are still too few. The idea is to try to change the parameter of numbers of generations.

In order to improve the classification accuracy rate, the number of generations was in-

creased 10 times to 100,000 to see what the results will be. Similar to the case of 10,000 generations, the experiment of 100,000 generations does 6-fold cross-validation as well. Figures B.1225 to B.1248 are the graphs of classification accuracies for 100,000 generations.

Figure B.1225: Cross-validation 1 for 100,000 generations training data except patterns 1, 2

Figure B.1226: Cross-validation 1 for 100,000 generations training data except patterns 3, 4

Figure B.1227: Cross-validation 1 for 100,000 generations training data except patterns 5, 6

Figure B.1228: Cross-validation 1 for 100,000 generations training data except patterns 7, 8

Figure B.1229: Cross-validation 1 for 100,000 generations training data except patterns 9, 10

Figure B.1230: Cross-validation 1 for 100,000 generations training data except patterns 11, 12

Figure B.1231: Cross-validation 1 for 100,000 generations testing data patterns 1, 2

Figure B.1232: Cross-validation 1 for 100,000 generations testing data patterns 3, 4

Figure B.1233: Cross-validation 1 for 100,000 generations testing data patterns 5, 6

Figure B.1234: Cross-validation 1 for 100,000 generations testing data patterns 7, 8

Figure B.1235: Cross-validation 1 for 100,000 generations testing data patterns 9, 10

Figure B.1236: Cross-validation 1 for 100,000 generations testing data patterns 11, 12

Figure B.1237: Cross-validation 2 for 100,000 generations training data except patterns 1, 12

Figure B.1238: Cross-validation 2 for 100,000 generations training data except patterns 2, 11

Figure B.1239: Cross-validation 2 for 100,000 generations training data except patterns 3, 10

Figure B.1240: Cross-validation 2 for 100,000 generations training data except patterns 4, 9

Figure B.1241: Cross-validation 2 for 100,000 generations training data except patterns 5, 8

Figure B.1242: Cross-validation 2 for 100,000 generations training data except patterns 6, 7

Figure B.1243: Cross-validation 2 for 100,000 generations testing data patterns 1, 12

Figure B.1244: Cross-validation 2 for 100,000 generations testing data patterns 2, 11

Figure B.1245: Cross-validation 2 for 100,000 generations testing data patterns 3, 10

Figure B.1246: Cross-validation 2 for 100,000 generations testing data patterns 4, 9

Figure B.1247: Cross-validation 2 for 100,000 generations testing data patterns 5, 8

Figure B.1248: Cross-validation 2 for 100,000 generations testing data patterns 6, 7

In order to create a wider variety of data, three more numbers of generations, 5,000 generations, 20,000 generations, and 50,000 generations, are set and the accuracy rates are compared. Here figures B.1249 to B.1272 illustrate the outcome of 5,000 generations.

Figure B.1249: Cross-validation 1 for 5,000 generations training data except patterns 1, 2

Figure B.1250: Cross-validation 1 for 5,000 generations training data except patterns 3, 4

Figure B.1251: Cross-validation 1 for 5,000 generations training data except patterns 5, 6

Figure B.1252: Cross-validation 1 for 5,000 generations training data except patterns 7, 8

Figure B.1253: Cross-validation 1 for 5,000 generations training data except patterns 9, 10

Figure B.1254: Cross-validation 1 for 5,000 generations training data except patterns 11, 12

Figure B.1255: Cross-validation 1 for 5,000 generations testing data patterns 1, 2

Figure B.1256: Cross-validation 1 for 5,000 generations testing data patterns 3, 4

Figure B.1257: Cross-validation 1 for 5,000 generations testing data patterns 5, 6

Figure B.1258: Cross-validation 1 for 5,000 generations testing data patterns 7, 8

Figure B.1259: Cross-validation 1 for 5,000 generations testing data patterns 9, 10

Figure B.1260: Cross-validation 1 for 5,000 generations testing data patterns 11, 12

Figure B.1261: Cross-validation 2 for 5,000 generations training data except patterns 1, 12

Figure B.1262: Cross-validation 2 for 5,000 generations training data except patterns 2, 11

Figure B.1263: Cross-validation 2 for 5,000 generations training data except patterns 3, 10

Figure B.1264: Cross-validation 2 for 5,000 generations training data except patterns 4, 9

Figure B.1265: Cross-validation 2 for 5,000 generations training data except patterns 5, 8

Figure B.1266: Cross-validation 2 for 5,000 generations training data except patterns 6, 7

Figure B.1267: Cross-validation 2 for 5,000 generations testing data patterns 1, 12

Figure B.1268: Cross-validation 2 for 5,000 generations testing data patterns 2, 11

Figure B.1269: Cross-validation 2 for 5,000 generations testing data patterns 3, 10

Figure B.1270: Cross-validation 2 for 5,000 generations testing data patterns 4, 9

Figure B.1271: Cross-validation 2 for 5,000 generations testing data patterns 5, 8

Figure B.1272: Cross-validation 2 for 5,000 generations testing data patterns 6, 7

Here is the end of the 5,000 generation experiments. Figures B.1273 to B.1296 display results of the 20,000 generations experiments.

Figure B.1273: Cross-validation 1 for 20,000 generations training data except patterns 1, 2

Figure B.1274: Cross-validation 1 for 20,000 generations training data except patterns 3, 4

Figure B.1275: Cross-validation 1 for 20,000 generations training data except patterns 5, 6

Figure B.1276: Cross-validation 1 for 20,000 generations training data except patterns 7, 8

Figure B.1277: Cross-validation 1 for 20,000 generations training data except patterns 9, 10

Figure B.1278: Cross-validation 1 for 20,000 generations training data except patterns 11, 12

Figure B.1279: Cross-validation 1 for 20,000 generations testing data patterns 1, 2

Figure B.1280: Cross-validation 1 for 20,000 generations testing data patterns 3, 4

Figure B.1281: Cross-validation 1 for 20,000 generations testing data patterns 5, 6

Figure B.1282: Cross-validation 1 for 20,000 generations testing data patterns 7, 8

Figure B.1283: Cross-validation 1 for 20,000 generations testing data patterns 9, 10

Figure B.1284: Cross-validation 1 for 20,000 generations testing data patterns 11, 12

Figure B.1285: Cross-validation 2 for 20,000 generations training data except patterns 1, 12

Figure B.1286: Cross-validation 2 for 20,000 generations training data except patterns 2, 11

Figure B.1287: Cross-validation 2 for 20,000 generations training data except patterns 3, 10

Figure B.1288: Cross-validation 2 for 20,000 generations training data except patterns 4, 9

Figure B.1289: Cross-validation 2 for 20,000 generations training data except patterns 5, 8

Figure B.1290: Cross-validation 2 for 20,000 generations training data except patterns 6, 7

Figure B.1291: Cross-validation 2 for 20,000 generations testing data patterns 1, 12

Figure B.1292: Cross-validation 2 for 20,000 generations testing data patterns 2, 11

Figure B.1293: Cross-validation 2 for 20,000 generations testing data patterns 3, 10

Figure B.1294: Cross-validation 2 for 20,000 generations testing data patterns 4, 9

Figure B.1295: Cross-validation 2 for 20,000 generations testing data patterns 5, 8

Figure B.1296: Cross-validation 2 for 20,000 generations testing data patterns 6, 7

Here is the end of the 20,000 generation experiments. Figures B.1297 to B.1320 display results of the 50,000 generations experiments.

Figure B.1297: Cross-validation 1 for 50,000 generations training data except patterns 1, 2

Figure B.1298: Cross-validation 1 for 50,000 generations training data except patterns 3, 4

Figure B.1299: Cross-validation 1 for 50,000 generations training data except patterns 5, 6

Figure B.1300: Cross-validation 1 for 50,000 generations training data except patterns 7, 8

Figure B.1301: Cross-validation 1 for 50,000 generations training data except patterns 9, 10

Figure B.1302: Cross-validation 1 for 50,000 generations training data except patterns 11, 12

Figure B.1303: Cross-validation 1 for 50,000 generations testing data patterns 1, 2

Figure B.1304: Cross-validation 1 for 50,000 generations testing data patterns 3, 4

Figure B.1305: Cross-validation 1 for 50,000 generations testing data patterns 5, 6

Figure B.1306: Cross-validation 1 for 50,000 generations testing data patterns 7, 8

Figure B.1307: Cross-validation 1 for 50,000 generations testing data patterns 9, 10

Figure B.1308: Cross-validation 1 for 50,000 generations testing data patterns 11, 12

Figure B.1309: Cross-validation 2 for 50,000 generations training data except patterns 1, 12

Figure B.1310: Cross-validation 2 for 50,000 generations training data except patterns 2, 11

Figure B.1311: Cross-validation 2 for 50,000 generations training data except patterns 3, 10

Figure B.1312: Cross-validation 2 for 50,000 generations training data except patterns 4, 9

Figure B.1313: Cross-validation 2 for 50,000 generations training data except patterns 5, 8

Figure B.1314: Cross-validation 2 for 50,000 generations training data except patterns 6, 7

Figure B.1315: Cross-validation 2 for 50,000 generations testing data patterns 1, 12

Figure B.1316: Cross-validation 2 for 50,000 generations testing data patterns 2, 11

Figure B.1317: Cross-validation 2 for 50,000 generations testing data patterns 3, 10

Figure B.1318: Cross-validation 2 for 50,000 generations testing data patterns 4, 9

Figure B.1319: Cross-validation 2 for 50,000 generations testing data patterns 5, 8

Figure B.1320: Cross-validation 2 for 50,000 generations testing data patterns 6, 7

Bibliography

Beyer, H.-G. [2007], 'Evolution Strategies'.

URL: *http://www.scholarpedia.org/article/Evolution strategies*

- Bhanu, B., Yu, J., Tan, X. & Lin, Y. [2004], Feature synthesis using genetic programming for face expression recognition, *in* 'in Proceedings of Genetic and Evolutionary Computation Conference', pp. 896–907.
- Bost, R., Popa, R. A., Tu, S. & Goldwasser, S. [2015], Machine Learning Classification over Encrypted Data, *in* '22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2014'. **URL:** *http://www.internetsociety.org/doc/machine-learning-classification-over-encrypteddata; http://dblp.uni-trier.de/rec/bib/conf/ndss/BostPTG15*
- Cagnoni, S. & Vanneschi, L. [2010], *Evolutionary Computation: A Brief Overview*, John Wiley & Sons, Ltd, pp. 3–15. **URL:** *http://dx.doi.org/10.1002/9780470973134.ch2*
- Chibelushi, C. C. & Bourel, F. [n.d.], 'Facial Expression Recognition: A Brief Tutorial Overview'. $\mathbf{URL: } http://homepages.inf.edu.c.uk/rbf/CVonline/LOCAL_COPIES/CHIBELUSHI1/CCC_{FB}$
- Corporation, N. I. [2015], 'The Different Parts of an FPGA'. **URL:** $http://sine.ni.com/np/app/main/p/ap/imc/lang/en/pg/1/sn/n17:imc,n19:FPGA_target/fm$
- Dawkins, R. & Bentley, P. J. [1999], *Evolutionary Design by computers*, Morgan Kaufmann Publishers, Inc. San Francisco, California, pp. 1–8.
- Dumitrescu, D., Lazzerini, B., Jain, L. G. & Dumitrescu, A. [2000*a*], *Evolutionary Computation*, CRC Press LLC, pp. 1–18.
- Dumitrescu, D., Lazzerini, B., Jain, L. G. & Dumitrescu, A. [2000*b*], *Evolutionary Computation*, CRC Press LLC, pp. 153–183.
- Eiben, A. E. & Smith, J. E. [2003*a*], *Introduction to Evolutionary Computing*, Springer Berlin / Heidelberg, pp. 1–7.
- Eiben, A. E. & Smith, J. E. [2003*b*], *Introduction to Evolutionary Computing*, Springer Berlin / Heidelberg, pp. 18–24.
- Eiben, A. E. & Smith, J. E. [2003*c*], *Introduction to Evolutionary Computing*, Springer Berlin / Heidelberg, pp. 15–18.
- Eiben, A. E. & Smith, J. E. [2003*d*], *Introduction to Evolutionary Computing*, Springer Berlin / Heidelberg, pp. 164–169.
- Eiben, A. E. & Smith, J. E. [2003*e*], *Introduction to Evolutionary Computing*, Springer Berlin / Heidelberg, pp. 190–196.
- Fogel, L., Owens, A. & Walsh, M. [1966], *Artificial Intelligence Through Simulated Evolution*, John Wiley & Sons. **URL:** *https://books.google.co.uk/books?id=QMLaAAAAMAAJ*
- Glette, K., Torresen, J., Gruber, T., Sick, B., Kaufmann, P. & Platzner, M. [2008], Comparing Evolvable Hardware to Conventional Classifiers for Electromyographic Prosthetic Hand Control, *in* 'Adaptive Hardware and Systems, 2008. AHS '08. NASA/ESA Conference on', pp. 32–39.
- Glette, K., Torresen, J., Kaufmann, P. & Platzner, M. [2008], A Comparison of Evolvable Hardware Architectures for Classification Tasks, *in* G. Hornby, L. Sekanina & P. Haddow, eds, 'Evolvable Systems: From Biology to Hardware', Springer Berlin / Heidelberg, pp. 22– 33.
- Glette, K., Torresen, J. & Yasunaga, M. [2007], An Online EHW Pattern Recognition System Applied to Face Image Recognition, *in* M. Giacobini, ed., 'Applications of Evolutionary Computing', Springer Berlin / Heidelberg, pp. 271–280.
- Goldberg, D. E. [1989], *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Company, Inc., pp. 1–18.
- Greenwood, G. W. & Tyrrell, A. M. [2007*a*], *Introduction to Evolvable Hardware*, John Wiley & Sons, Inc., Hoboken, New Jersey, p. 17.
- Greenwood, G. W. & Tyrrell, A. M. [2007*b*], *Introduction to Evolvable Hardware*, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 18–29.
- Greenwood, G. W. & Tyrrell, A. M. [2007*c*], *Introduction to Evolvable Hardware*, John Wiley & Sons, Inc., Hoboken, New Jersey, p. 31.
- Greenwood, G. W. & Tyrrell, A. M. [2007*d*], *Introduction to Evolvable Hardware*, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 1–15.
- Greenwood, G. W. & Tyrrell, A. M. [2007*e*], *Introduction to Evolvable Hardware*, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 95–96.
- Greenwood, G. W. & Tyrrell, A. M. [2007*f*], *Introduction to Evolvable Hardware*, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 35–93.
- Hansen, N., Arnold, D. V. & Auger, A. [2013], 'Evolution Strategies'.
- Holland, J. H. [1973], 'Genetic Algorithms and the Optimal Allocation of Trials', *SIAM Journal on Computing* **2**(2), 88–105. **URL:** *http://dx.doi.org/10.1137/0202009*
- Holland, J. H. [1992*a*], *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, MA, USA.
- Holland, J. H. [1992*b*], *Adaptation in Natural and Artificial Systems*, First MIT Press edition, pp. 66–74.
- Jong, K. A. D. [2006*a*], *Evolutionary Computation*, The MIT Press, pp. 223–224.
- Jong, K. A. D. [2006*b*], *Evolutionary Computation*, The MIT Press, pp. 103–104.
- Koza, J. R. [1992*a*], *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press.
- Koza, J. R. [1992*b*], *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, pp. 62–119.
- Koza, J. R. [1994], *Genetic Programming II*, The MIT Press.
- Lien, J.-J. J., Kanade, T., Cohn, J. & Li, C.-C. [1998], Automated Facial Expression Recognition Based on FACS Action Units, *in* 'Third IEEE International Conference on Automatic Face and Gesture Recognition', pp. 390–395.

Malcolm Heywood [2012], 'Holland's GA Schema Theorem'.

URL: *http://www.cse.unr.edu/ sushil/class/gas/notes/GASchemaTheorem2.pdf*

Martin, P. [2003], Genetic Programming in Hardware, PhD thesis, University of Essex.

MIBI [n.d.], 'Genetic Algorithm'. **URL:** *http://140.124.77.80/intro GA.htm*

- Miller, J. F. [2011], *Cartesian Genetic Programming*, Springer-Verlag Berlin / Heidelberg.
- Miller, J. & Tempesti, G. [2011], Bio-inspired Computing (Cartesian Genetic Programming). Lecture.
- Miller, J. & Thomson, P. [2000], Cartesian Genetic Programming, *in* R. Poli, W. Banzhaf, W. Langdon, J. Miller, P. Nordin & T. Fogarty, eds, 'Genetic Programming', Vol. 1802 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 121–132.
- Mitchell, M. [1998], *An Introduction to Genetic Algorithms*, The MIT Press, pp. 8–10.
- Müller, S. [2002], Bio-inspired Optimization Algorithms for Engineering Applications, PhD thesis, Swiss Federal Institute of Technology ZÜRICH.
- Obitko, M. [1998], 'Genetic Algorithm'.

URL: *http://www.obitko.com/tutorials/genetic-algorithms/ga-basic-description.php*

- Rechenberg, I. [1973], *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Problemata, 15, Frommann-Holzboog. **URL:** *https://books.google.co.uk/books?id=-WAQAQAAMAAJ*
- Schapire, R. [n.d.], 'Machine Learning Algorithms for Classification'. **URL:** *http://www.cs.princeton.edu/ schapire/talks/picasso-minicourse.pdf*
- Schwefel, H. [1995], *Evolution and Optimum Seeking*, Sixth Generation Computer Technologies, Wiley. **URL:** *https://books.google.co.uk/books?id=dfNQAAAAMAAJ*

- Sekanina, L. [2004*a*], *Evolvable Components*, Springer-Verlag Berlin / Heidelberg, pp. 34–36.
- Sekanina, L. [2004*b*], *Evolvable Components*, Springer-Verlag Berlin / Heidelberg, pp. 115– 117.
- Sekanina, L. [2004*c*], *Evolvable Components*, Springer-Verlag Berlin / Heidelberg, pp. 49–51.
- Teller, A. & Veloso, M. [1995], Algorithm Evolution for Face Recognition: What Makes a Picture Difficult, *in* 'International Conference on Evolutionary Computation', IEEE Press, Perth, Australia, pp. 608–613.
- Torresen, J. [2008], 'Adapting Systems by Evolving Hardwarre'. **URL:** *http://web.stanford.edu/class/ee380/Abstracts/080123^T orresen.pdf*
- Torresen, J., Bakke, J. W. & Sekanina, L. [2004], Recognizing speed limit sign numbers by evolvable hardware, *in* 'In Proc. of Parallel Problem Solving from Nature VIII (PPSN VIII), Lecture Notes in Computer Science', Springer-Verlag, pp. 682–691.
- Torresen, J., Senland, G. & Glette, K. [2008], Partial Reconfiguration Applied in an On-line Evolvable Pattern Recognition System, *in* 'NORCHIP, 2008.', pp. 61–64.
- Trefzer, M. [2011], Evolvable Logic on FPGA. The Lab of the lecture "Introduction to Evolvable Hardware".
- Vellasco, M. M. B., Zebulum, R. S. & Pacheco, M. A. [2001], *Evolutionary Electronics: Automatic Design of Electronic Circuits and Systems by Genetic Algorithms*, 1st edn, CRC Press, Inc., Boca Raton, FL, USA.
- Wall, M. [n.d.], 'Introduction to Genetic Algorithms'. **URL:** *http://lancet.mit.edu/mbwall/presentations/IntroToGAs*
- Wang, J., Jung, J., Lee, Y.-m. & Lee, C. [2005], Using Reconfigurable Architecture-Based Intrinsic Incremental Evolution to Evolve a Character Classification System, *in* Y. Hao, J. Liu, Y. Wang, Y.-m. Cheung, H. Yin, L. Jiao, J. Ma & Y.-C. Jiao, eds, 'Computational Intelligence and Security', Springer Berlin / Heidelberg, pp. 216–223.
- Widiger, H., Salomon, R. & Timmermann, D. [2006], Packet Classification with Evolvable Hardware Hash Functions – An Intrinsic Approach, *in* A. Ijspeert, T. Masuzawa & S. Kusumoto, eds, 'Biologically Inspired Approaches to Advanced Information Technology', Springer Berlin / Heidelberg, pp. 64–79.
- Wolf, W. [2004], *FPGA-based System Design*, Prentice-Hall Professional Technical Reference, pp. 105–162.

Xilinx XUPV5-LX110T Evalution Platform [n.d.].

XUPV2P UserGuide v1.2 [n.d.].

- Yasunaga, M., Nakamura, T., Yoshihara, I. & Kim, J. [2000], Genetic Algorithm-Based Design Methodology for Pattern Recognition Hardware, *in* 'Evolvable Systems: From Biology to Hardware', Springer Berlin / Heidelberg, pp. 264–273.
- Zebulum, R. S., Stoica, A. & Keymeulen, D. [2000], *A flexible model of a CMOS field programmable transistor array targeted for hardware evolution*, Springer.
- Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P. N. & Zhang, Q. [2011], 'Multiobjective evolutionary algorithms: A survey of the state of the art', *Swarm and Evolutionary Computation* **1**(1), 32–49. **URL:** *http://www.sciencedirect.com/science/article/pii/S2210650211000058*
- Zitzler, E., Laumanns, M. & Bleuler, S. [2003], A Tutorial on Evolutionary Multiobjective Optimization, *in* X. Gandibleux, M. Sevaux, K. Sörensen & V. T'kindt, eds, 'Metaheuristics for Multiobjective Optimisation', Vol. 535 of *Lecture Notes in Economics and Mathematical Systems*, Springer Berlin Heidelberg, pp. 3–37. **URL:** *http://dx.doi.org/10.1007/978-3-642-17144-4*¹
- Zitzler, E., Laumanns, M. & Thiele, L. [2001], SPEA2: Improving the Strength Pareto Evolutionary Algorithm, Technical Report TIK-Report 103, Computer Engineering and Networks Laboratory (TIK), Department of Electrical Engineering, Swiss Federal Institute of Technology (ETH) Zurich.
- Zitzler, E. & Thiele, L. [1999], 'Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach', *Evolutionary Computation, IEEE Transactions on* **3**(4), 257–271.
- Zitzler Eckart, Deb Kalyanmoy & Thiele Lothar [2000], 'Comparison of Multiobjective Evolutionary Algorithms: Empirical Results', *Evolutionary Computation* **8**(2), 173–195. doi: 10.1162/106365600568202.