The Application of Evolutionary Algorithms towards the Diagnosis of Parkinson's Disease

Stephen L. Smith
The University of York
Department of Electronics
Heslington, York YO10 5DD, UK

sls5@ohm.york.ac.uk

Patrick Gaughan
The University of York
Department of Electronics
Heslington, York YO10 5DD, UK

David M. Halliday
The University of York
Department of Electronics
Heslington, York YO10 5DD, UK

dh20@ohm.york.ac.uk

Quan Ju
The University of York
Department of Electronics
Heslington, York YO10 5DD, UK

Nabil M. Aly
University Hospital Aintree
Lower Lane
Liverpool L0 7AL, UK

nmaly@doctors.org.uk

Jeremy R. Playfer
Royal Liverpool and Broadgreen
University Hospitals
Liverpool, UK

jrplayfer@doctors.org.uk

ABSTRACT
This paper describes the novel application of an evolutionary algorithm to discriminate Parkinson's patients from age-matched controls in their response to simple figure-copying tasks. The reliable diagnosis of Parkinson's disease is notoriously difficult to achieve with misdiagnosis reported to be as high as 25% of cases. The approach described in this paper aims to distinguish between the velocity profiles of pen movements of patients and controls to identify distinguishing artifacts that may be indicative of the Parkinson's symptom bradykinesia. Results are presented for 12 patients with Parkinson's disease and 10 age-match controls. An algorithm was evolved using half the patient and age-matched control responses, which was then successfully used to correctly classify the remaining responses.

Categories and Subject Descriptors
I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*heuristic methods*;
I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*medicine and science*

GECCO '06, July 8–12, 2004, Seattle, Washinton, USA.

General Terms
Algorithms, Measurement, Experimentation.

Keywords
Parkinson's disease, evolutionary algorithms, Cartesian genetic programming.

1. INTRODUCTION
Parkinson Disease (PD) is a common, chronic, progressive neurodegenerative brain disease, afflicting about 1 person in 1000, and about 1 person in 100 over the age of 60. The disease sets in insidiously, and in most patients progresses relentlessly, on average within 10 years, to a state of total physical incapacitation. The symptoms usually start on one side of the body (hemiparkinsonism) but later spread to the other side. About 80% of patients suffer from idiopathic Parkinson disease for which no cause is known. However, the diagnosis of idiopathic PD is based on clinical features which can have poor sensitivity with about 25% of patients diagnosed with the disease actually having other conditions [11]. Considerable research is being conducted to improve the diagnosis of the condition, but most studies to date are reliant on laboratory-based experimentation.

The authors are concerned with developing a non-invasive computer-based test that can be conducted in the clinical environment and the doctor's surgery, using commonly available computing peripherals. Work to date by the authors on such a computer-based assessment using figure-copying tasks has generated promising results. However, it is clear the identification

and quantification of the symptoms of Parkinson's disease is a non-trivial problem that is highly dependent on many factors relating to the patient's physical as well as medical condition.

This paper reports the application of an evolutionary algorithm to classify patients' responses to such a figure copying task with far greater accuracy than has previously been possible.

Section 2 describes the authors' previous approach to the assessment of Parkinson's disease using a conventional figure copying task, Section 3 describes the implementation of the evolutionary algorithm. Results for Parkinson's patients and age matched controls are presented in Section 4. Conclusions and further work are then discussed in Section 5.

2. COMPUTER-BASED ASSESSMENT OF PARKINSONS DISEASE

Parkinson Disease has three main symptoms: *Tremor, Rigidity* and *Bradykinesia*:

Parkinsonian tremor is an involuntary rhythmically alternating contraction of agonist and antagonist muscles, and is especially prominent at rest, and decreases, or ceases, during active movement. The frequency of this tremor usually lies between 3–8 cycles/sec (Hz) and is most evident, and often starts first, in the extremities (fingers, hands, legs), but the head, lips and tongue are usually also affected.

The rigidity of Parkinson disease is a stiffness of the skeletal muscles and joints due to increased muscle tone. It is revealed by the presence of a rhythmically jerking ('cogwheel') resistance offered by an extremity to passive movement. Rigidity can affect all the musculature of the body.

Bradykinesia is the core disabling feature of Parkinson disease. It consists of difficulty, slowness (bradykinesia proper) or virtual inability (akinesia) in initiating and executing movements or modifying ongoing motor activity. Poverty of spontaneous movement (hypokinesia), loss of normal associated movements, masked facial expression and sudden 'freezing' in the middle of a motor performance are all part of the disturbance.

The aim of this work is to devise a computer-based system that is capable of measuring these symptoms to aid diagnosis and inform administration of medication.

2.1 Data acquisition

The computer-based assessment comprises two parts: data acquisition and data processing. Data acquisition stage is the digitization of the patient's drawing in attempting a conventional figure-copying task, termed here, the task domain. Once the patient's response has been acquired in digital form, data processing is applied to extract and quantify the symptoms of interest, in this case tremor and bradykinesia.

One of the aims of this work is to preserve, as far as possible, a conventional writing environment. This will avoid unnecessary distress to the patient, while allowing comparison with other traditional tests. To help achieve this aim, a commercially available digitizing tablet with stylus interface was employed. The study described here used a Wacom tablet, which has the advantage of utilizing a stylus with ball point refills, that can be used in the same way as conventional ball-point pen on standard paper, thus reproducing a conventional "pen and paper" environment (Figure 1).

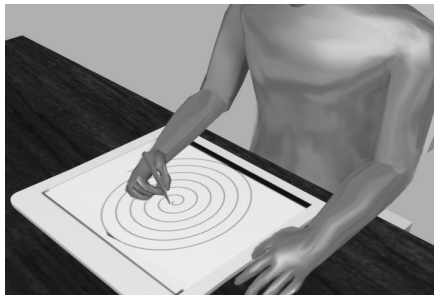


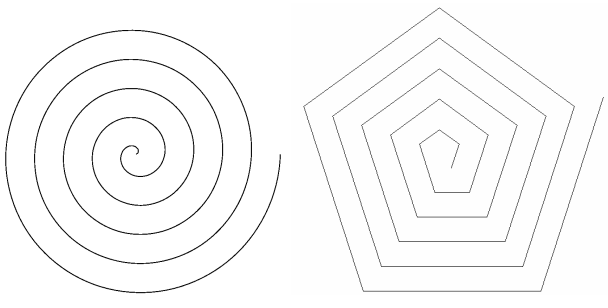
Figure 1. Figure copying task using a conventional digitizing tablet.

2.2 Task domain

The patient's response to the task presented is a drawing activity that is digitized in real-time as a set of x-y co-ordinate pairs giving information about pen position. In addition, information regarding pen pressure and pen tilt is provided. The present study uses only the positional information.

The task domain is particular to the test being conducted and hence the neurological condition of interest. The object of a task domain is to accentuate those symptoms of the neurological disorder to be evaluated. The figure required to be copied by the patient is placed on the digitizing tablet in printed form and covered with a sheet of tracing paper on which the patient traces a copy. The experimental protocol required patients to trace the shape as fast as they could. Patients were asked to attempt the figure-copying task at least three times, although this was often dictated by the condition of the individual.

The geometric shape presented for the patient to copy in this study is based on the Archimedes spiral (Figure 2a), which is commonly used in assessing tremor in Parkinson's patients [2]. To make the shape useful for assessing bradykinesia, it was modified, replacing the smooth spiral with pentagon-like edges as shown in Figure 2b. The justification for making this modification is that by introducing a sequential aspect, the change from drawing one edge to another will provide a focus at which symptoms of bradykinesia may be observed [1].



(a) (b)

Figure 2. Archimedes spiral (a) as inspiration for the task domain shape to be copied (b)

2.3 Measurement of symptoms

The detection and measurement of tremor is well documented and can be a useful indicator of Parkinson's disease but is by no means a unique symptom as it is only found in a proportion of cases and can also be confused with other neurological dysfunctions. Similarly, rigidity is not a unique symptom and is very difficult to measure with conventional computer interfaces.

Although the term bradykinesia is defined as the slowness of a performed movement [1], it is commonly used synonymously with akinesia and hypokinesia, an expression of freezing and smaller movements respectively. To quantify these symptoms a measure of movement time and particularly, the movement velocity of the patient's pen, is required.

As the patient's drawing activity is digitized in real-time and at regular intervals, it is possible to determine the velocity of the pen at any instant. This can be achieved by calculating the distance between two coordinate positions and dividing this by the difference of the relative timestamps as shown in (1).

$$v_{ij} = \frac{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}{t_j - t_i} \qquad (1)$$

Where:

v_{ij} velocity of the pen between coordinates x_i y_i and x_j y_j

t_i , t_j respective times at which pen coordinates x_i y_i and x_j y_j were recorded

The above equation was used to provide a representation of the velocity of the pen for the duration of the patient's response to the task attempted. This has been termed as the *velocity profile* of the patient and an example is given in Figure 3. As might be expected, the velocity profile can reflect the nature of the task, e.g. in this case an acceleration and deceleration can often be associated with each side of the spiral pentagon.

The main thrust of data analysis to date has focused on the patient's pen velocity at the end of one edge of the spiral-pentagon and the beginning of the next. It is here that evidence of a slowing and of hesitation, commonly recognized as bradykinesia, is sought that will differentiate a Parkinson's patient from a normal control [1]. Subjective examination of velocity profiles obtained from the 12 Parkinson's patients and 10 controls was made with a view to identifying features that could be used as evidence of bradykinesia.

One candidate for such a feature is a two-part artifact, which is illustrated in Figure 4. The initial acceleration of the pen is described by the section of the velocity profile labeled "A". After a period of remission, a second period of acceleration greater than that of "A" is observed (labeled "B"). A total of 10 occurrences of the two-part velocity feature were detected in the responses of five separate patients. The feature was only located in one of the age-matched control drawing responses.

2.4 Summary of previous work

The feature described in Section 2.3 is one example of the type of analysis that can be applied to digitized pen movements over a variety of tasks, each of which can be designed to emphasize and quantify the particular symptom of interest. However, it is clear that identification of these features is a difficult and time consuming task and that automating this process using an evolutionary algorithm would be potentially beneficial.

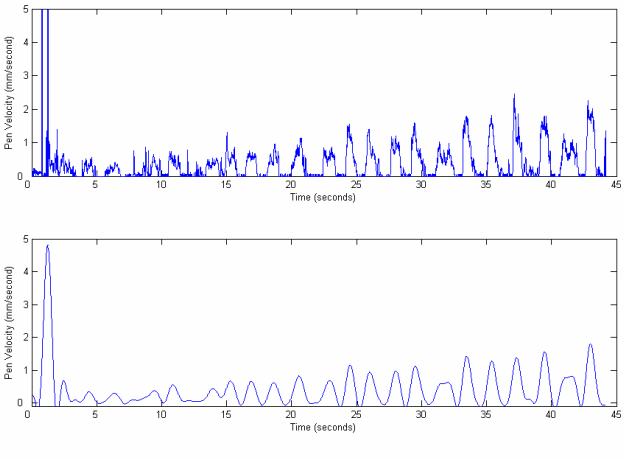


Figure 3. Example velocity profile of a patient's response (top) and after smoothing (bottom).

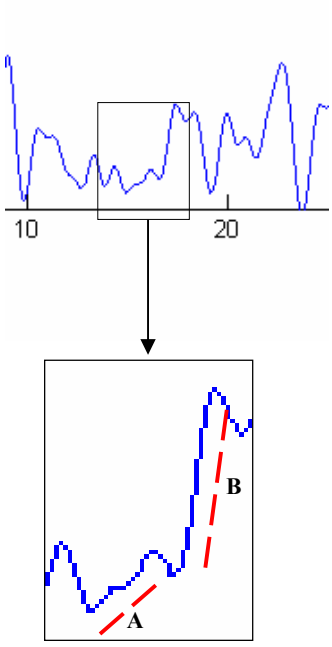


Figure 4. Example of two-part velocity feature investigated as evidence of bradykinesia.

3. APPLICATION OF EVOLUTIONARY ALGORITHM

The application of an evolutionary algorithm to the problem of locating features within a patient's velocity profiles requires the following preparations to be undertaken: the appropriate representation of the patient's data; customization of the chosen evolutionary algorithm; arrangement of patients' and age matched controls' responses in to suitable training and testing sets.

3.1 Representation of data

First of all, the velocity profile from each patient's response to the figure copying task needs to be converted in to a form that will be suitable for manipulation by the chosen evolutionary algorithm.

As described in Section 2.2, the relative velocities of the patient's pen, describing a two-part acceleration, is of particular interest in detection of the bradykinesia symptom of Parkinson's disease. For this reason, the acceleration of the pen through the duration of the patient's velocity profile is simply calculated by differentiating with respect to time. The calculated acceleration or gradient values are then quantized and encoded according to the rules described in Table 1. The data is now in an appropriate form for the chosen evolutionary algorithm.

Table 1. Quantization and encoding of patient's data.

Gradient range	Gradient encoding
gradient > 2	6
1 <= gradient <= 2	5
0 < gradient < 1	4
gradient == 0	3
-1 < gradient < 0	2
-2 <= gradient <= -1	1
gradient < -2	0

Where: gradient is the acceleration of the patient's pen.

3.2 Evolutionary algorithm

An implicit context representation of a Cartesian Genetic Program (CGP) was used for the evolutionary algorithm in this application.

Cartesian Genetic Programming (CGP) was first proposed by Miller [9,10] as an alternative representation for genetic programming which does not require the use of a parse-tree based programming language and does not exhibit uncontrolled expansion commonly termed bloat [3]. As opposed to the rigid tree structure representation of traditional GP, CGP permits the arrangement of functions in a far more flexible, typically rectangular format, referenced by conventional Cartesian co-ordinates.

A criticism of CGP (and GP in general) is that the location of genes within the chromosome has a direct or indirect influence on the resulting phenotype [7]. In other words, the order in which specific information regarding the definition of the GP is stored has a direct or indirect effect on the operation, performance and characteristics of the resulting program. Such effects are considered undesirable as they may mask or modify the role of the specific genes in the generation of the phenotype (or resulting program). Consequently, GPs are often referred to as possessing a direct or indirect context representation.

An alternative representation for GPs in which genes do not express positional dependence has been proposed by Lones and Tyrrell [6-8]. Termed implicit context representation, the order in which genes are used to describe the phenotype (or resulting program) is determined after their self-organized binding, based on their own characteristics and not their specific location within the genotype. The result is an implicit con-text representation version of traditional parse-tree based GP termed Enzyme Genetic Programming. The authors have since implemented an implicit context representation of CGP, termed Implicit Context Representation Cartesian Genetic Programming (IRCGP), specifically for the evolution of image processing filters [12].

Implicit context representation employs an enzyme model comprising a shape, activity and specificities (or binding sites) [4], as shown in Figure 5. Along with inputs and outputs, the enzyme model can be considered a program component, executing one of the functions listed in Table 2, from which a genetic program may be constructed. The shape describes how the enzyme is seen by other program components. Similarly, the binding sites determine the shape (and hence type) of program component the enzyme wishes to bind to. Finally, the activity determines the logical function the enzyme is to perform. A typical IRCGP will comprise a set number of inputs and outputs and a number of enzyme models or components. Initial values for each component's binding sites and logical function are assigned non-deterministically; the component's shape, however, is derived from a combination of its binding sites' shapes and logical function.

Once initialized, components are bound together to form a network, as shown in Figure 6. The order in which components are bound is determined by the closeness of match between a component's binding site shape and another component's shape. The best matching components are bound first and the process is repeated until a network has formed in which no further binding is possible.

Over time, components may evolve through mutation. Mutation is applied to the component's binding sites and logical function with a pre-determined probability. When this occurs, a new component shape is derived accordingly and may lead to different binding between components occurring. This in turn may result in a modified network.

The network of processing elements is arranged in 10 rows and 3 columns as shown in Figure 6. In addition, 10 input components and one output component can also be seen. The 10 input components are fed by 10 consecutive gradient data from the data described in Section 3.1. The value obtained at the output component is used to indicate whether a particular artifact (representing a Parkinson feature) is present.

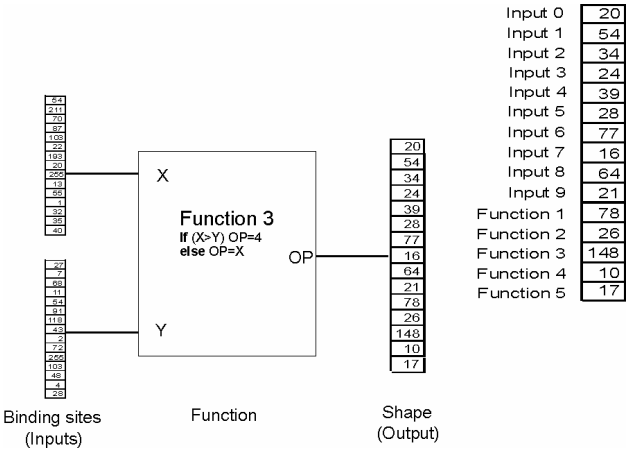


Figure 5. Example of a processing element that forms the evolutionary network.

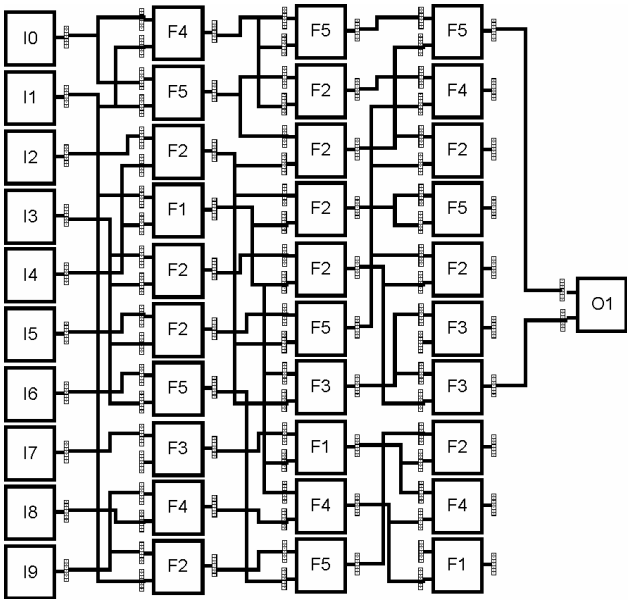


Figure 6. Example evolutionary network.

Table 2. Functions available for processing elements.

Function Index	Function Definition
F1	**if** (X>Y+3) OP=6 **else** OP=X
F2	**if** (X<Y+3) OP=0 **else** OP=Y
F3	**if** (X>Y) OP=4 **else** OP=X
F4	**if** (X<Y) OP=2 **else** OP=Y
F5	OP = (X+Y)/2

Where: X is the first input value to the component
Y is the second input value to the component
OP is the output value of the component

3.3 Fitness function

The fitness function is based on the desire to identify some artifact in the patient responses, but not in age-matched controls. The presence of the artifact is determined by a value returned by the output component of each individual network as being greater than 3; a value less or equal to 3 indicating the non-presence of the artifact. (The number 3 is the middle of the range of values possible at the output component.)

The fitness score comprises two parts: each dependent on whether a patient or age-matched control is being tested. For a Parkinson's patient response, the fitness score is the number of artifacts detected; conversely, in an age-matched controls response, the inverse is the case.

To achieve the aim of identifying a symptom of Parkinson's disease it is only necessary to detect one such artifact in a patient response. However, it is equally essential that no such artifacts are found in the age-matched controls. For this reason the fitness function used to evolve the network is weighted heavily in favor of non-detection of artifacts in the age-matched control responses. This is achieved by using an exponential function to bias the fitness scores accordingly.

3.4 Evolutionary parameters

For the results presented in this paper a network was evolved using a population size of 5 over 10000 generations. A conventional elitist strategy was adopted with mutation rate of 6% for the function used by each component and 3% for each dimension of the binding sites' shapes.

4. RESULTS

4.1 Patient population

12 patients with idiopathic PD (5/12, 42% female) were assessed as well as 10 controls (4/10, 40% female) who did not have PD or other neurological disorders, including stroke. Participants were enrolled from a PD specialist clinic and Day Hospital, after giving an informed consent approved by Liverpool Research Ethics Committee. The average age of PD patients enrolled in this study was 74.1 years (SD=8.4 years), and the exclusion criteria included drug-induced parkinsonism, Parkinson-plus and multisystem atrophy syndromes, Alzheimer's disease and significant cognitive impairment. The majority of the controls were relatives attending with patients at the Day hospital, but a few were patients attending the Day hospital for general rehabilitation or assessment, and the average age was 73.2 years (SD=5.3 years). In order to assess the performance of the system under conditions normally found in out-patient clinics, patients were not given any specific instructions regarding medication, and were tested under their normal medication regime.

The patient and age-matched control responses were arbitrarily split in to a training set and a testing set of approximately equal sizes.

4.2 Evolution Stage

Using the evolutionary algorithm described in Section 3, the evolution of a network that would discriminate between Parkinson's patient responses and age-matched controls was attempted. After 101 generations a network with control fitness of 99.068% and a patient fitness of 0.168% was evolved. It should not be surprising that the age-matched control fitness was not 100% as many of the patients would have been under the influence of medication when tested, compensating for some of the symptoms of Parkinson's disease. Equally, as described in Section 3.3 it is not of great concern that the patient fitness was low.

4.3 Testing Stage

The chromosome representing the network with the highest fitness was saved and used in the testing phase. Specifically, the evolved network was used to discriminate between 11 patient responses and 19 control responses which were not included in the evolution stage.

The results are shown in Figure 7. For each response the number of occurrences of the artifact identified by the evolved network is shown. Ideally, no artifacts should be present in the age-matched control responses and at least one occurrence of the artifact should be present in every patient response. As can be seen, artifacts have been located in every response, but importantly, more have

been located in patient responses than in the age-matched controls. More specifically, the age-matched controls all have five or less occurrences of the artifact, whereas the patient responses each have six or more occurrences of the artifact.

Response ID	Response type	Number of artifacts detected
C1	control response	1
C2	control response	2
C3	control response	2
C4	control response	3
C5	control response	2
C6	control response	2
C7	control response	1
C8	control response	4
C9	control response	3
C10	control response	2
C11	control response	2
C12	control response	2
C13	control response	1
C14	control response	3
C15	control response	1
C16	control response	2
C17	control response	4
C18	control response	4
C19	control response	5
P1	patient response	7
P2	patient response	7
P3	patient response	6
P4	patient response	8
P5	patient response	6
P6	patient response	7
P7	patient response	7
P8	patient response	8
P9	patient response	8
P10	patient response	10
P11	patient response	8

Figure 7. Results of testing the evolved algorithm with patient and age-matched control responses.

5. CONCLUSIONS

This paper has described the initial results of an evolutionary algorithm used to discriminate between Parkinson's patients' and age-matched controls' responses to a simple figure copying test. Although the discrimination between the two populations is not great, the authors believe that the results have demonstrated the potential in applying evolutionary algorithms to medical problems of this type, Further work is currently under way to fully characterize the evolution of these networks

6. ACKNOWLEDGMENTS

We would like to thanks the staff of Broadgreen Hospital, Liverpool, UK for their help in acquiring the patient data for this work,

7. REFERENCES

[1] Berardelli et al. *Brain*, 2001, 124 (11) 2131-2146

[2] Elble, R.J. & Koller, J. *Tremor*, John Hopkins, Baltimore, 1990.

[3] Langdon, W. Quadratic bloat in genetic programming. *In D. Whitley, D. Goldberg, and E. Cantu-Paz, editors, Proceedings of the 2000 Genetic and Evolutionary Computation Conference,* 2000, 451-458.

[4] Lones, M. A. and Tyrrell, A. M. Enzyme genetic programming. *Proc. 2001 Congress on Evolutionary Computation, J.-H. Kim, B.-T. Zhang, G. Fogel, and I. Kuscu (eds.), IEEE Press.* Vol. 2, 2001, 1183–1190.

[5] Lones, M. A. and Tyrrell, A. M. Crossover and Bloat in the Functionality Model of Enzyme Genetic Programming. *Proc. Congress on Evolutionary Computation,* 2002, 986-992.

[6] Lones, M. A. and Tyrrell, A. M. Biomimetic Representation with Enzyme Genetic Programming. *Journal of Genetic Programming and Evolvable Machines.* Vol.3 No.2, 2002, 193-217.

[7] Lones, M. A. *Enzyme Genetic Programming.* PhD Thesis, University of York, UK, 2003.

[8] Lones, M. A. and Tyrrell, A. M. Modelling biological evolvability: implicit context and variation filtering in enzyme generic programming. *BioSystems,* 2004.

[9] Miller, J.F. and Thomson, P. Cartesian genetic programming. *In Third European Conf. Genetic Programming, R. Poli, W. Banzhaf, W. B. Langdon, J. F. Miller, P. Nordin, and T. C. Fogarty (eds.),* LNCS 1802, 2000.

[10] Miller, J.F., Job, D. and Vasilev, V. K. Principles in the evolutionary design of digital circuits—Part I. *Genetic Programming and Evolvable Machines,* Vol. 1, 2000, 7–36.

[11] Playfer, JR. Parkinson's disease: Classic disease revisited. *Postgrad Med J.*, 73, 1997, 257-64.

[12] Smith, S.L., Leggett, S. and Tyrrell, A.M. An Implicit Context Representation for Evolving Image Processing Filters. *Proceedings of the 7th Workshop on Evolutionary Computation in Image Analysis and Signal Processing (EvoIASP'05),* LNCS 3449. 2005, 407-416.