Virtual Reality Visual Data Mining via Neural Networks Obtained from Multi-objective Evolutionary Optimization: Application to Geophysical Prospecting
Valdés, Julio; Barton, Alan

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version.

Publisher's version / Version de l'éditeur:

2006 IEEE International Joint Conference on Neural Networks (IJCNN 2006) [Proceedings], 2006

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=ac150ffc-f6be-453e-8434-97be3579af9d
https://publications-cnrc.canada.ca/fra/voir/objet/?id=ac150ffc-f6be-453e-8434-97be3579af9d

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L’accès à ce site Web et l’utilisation de son contenu sont assujettis aux conditions présentées dans le site https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D’UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n’arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.
Virtual Reality Visual Data Mining via Neural Networks obtained from Multi-objective Evolutionary Optimization: Application to Geophysical Prospecting *

Valdés, J., and Barton, A.
July 2006

Abstract—A method for the construction of Virtual Reality spaces for virtual data mining using multi-objective optimization with genetic algorithms on non-linear discriminant (NDA) neural networks is presented. Two neural network layers (output and last hidden) are used for the construction of simultaneous solutions for: a supervised classification of data patterns and an unsupervised similarity structure preservation between the original data matrix and its image in the new space. A set of spaces are constructed from selected solutions along the Pareto front. This strategy represents a conceptual improvement over spaces computed by single-objective optimization. In addition, genetic programming (in particular gene expression programming) is used for finding analytic representations of the complex mappings generating the spaces (a composition of NDA and orthogonal principal components). The presented approach is domain independent and is illustrated via application to the geophysical prospecting of caves.

I. INTRODUCTION

Increasing data generation rates, data kinds (relational, graphic, symbolic, etc.) and pattern relationships (geometrical, logical, etc.) require the development of procedures facilitating more rapid and intuitive understanding of inherent data structure. Moreover, the increasing complexity of data analysis makes it more difficult for a user (not necessarily a mathematician or data mining expert), to extract useful information out of results generated by the various techniques. This makes visual representation directly appealing.

The purpose of this paper is to explore construction of high quality VR spaces for visual data mining through the use of multi-objective optimization based on genetic algorithms (MOGA) operating on non-linear discriminant (NDA) neural networks. Both the NDA network output and the output of the last hidden layer are used for constructing solutions that simultaneously satisfy: class separability, and similarity structure preservation. Thus, a set of spaces can be obtained in which the different objectives are expressed to different degrees; with the proviso that no other spaces could improve any of the considered criteria individually (if spaces are selected from the Pareto front). This strategy represents a conceptual improvement over spaces that have been computed from the solutions obtained by single-objective optimization algorithms in which the objective function is a weighted composition involving different criteria. In addition, genetic programming, in particular gene expression programming (GEP), is applied with the purpose of finding explicit analytic representations of the complex mappings generating the spaces (a composition of NDA and orthogonal principal components transformation).

This approach is applied to a geophysical prospecting problem: namely, the detection of underground caves.

II. VIRTUAL REALITY REPRESENTATION OF RELATIONAL STRUCTURES AND VISUAL DATA MINING

A visual virtual reality based data mining technique extending the concept of 3D modeling to relational structures was presented in [23], [25], (see also http://www.hybriddstrategies.com). It is oriented to the understanding of large heterogeneous, incomplete and imprecise data, as well as other forms of structured and unstructured knowledge. In this approach, the data objects are considered as tuples from a heterogeneous space [24] Fig.1. A heterogeneous domain is a Cartesian product of a collection of source sets: \( \Psi^\eta = \Psi_1 \times \cdots \times \Psi_n \), where \( n > 0 \) is the number of information sources to consider.

<table>
<thead>
<tr>
<th>Nominal</th>
<th>Ordinal</th>
<th>Ratio</th>
<th>Fuzzy</th>
<th>Image</th>
<th>Signal</th>
<th>Graph</th>
<th>Doc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>high</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>?</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blue</td>
<td>low</td>
<td>-7.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. An example of a heterogeneous database. Nominal, ordinal, ratio, fuzzy, image, signal, graph, and document data are mixed. The symbol ? denotes a missing value.

A virtual reality space is the tuple \( \Upsilon = \langle Q, G, B, \mathbb{R}^r, g_o, l, g_r, b, r \rangle \), where \( Q \) is a relational structure (\( Q = \langle O, \Gamma^r \rangle \), \( O \) is a finite set of objects, and \( \Gamma^r \) is a set of relations); \( G \) is a non-empty set of geometries representing the different objects and relations; \( B \) is a non-empty set of behaviors of the objects in the virtual world;
\( \mathbb{R}^m \subset \mathbb{R}^m \) is a metric space of dimension \( m \) (euclidean or not) which will be the actual virtual reality geometric space. The other elements are mappings: \( g_o : O \to G, l : O \to \mathbb{R}^m, g_r : \Gamma \to G, b : O \to B. \)

Of particular importance is the mapping \( l \), where several desiderata can be considered for building a VR-space. From a preservation criteria \([15], [3], [1]\), or minimize some measure \([14], [25] \). From an unsupervised perspective, the role of \( O \) implicit obtained using approaches of this kind are only Levenberg-Marquardt, and others. However, they suffer from images of objects in the new space (the virtual reality space) between the multiple optimization (local search) techniques. The classical algorithm requires each individual to have one structure the next population in a series of evolving populations. A genetic algorithm permits particular sequences of operations. Networks suffer from local entrapment. This problem was approached in \([27]\) with hybrid stochastic-deterministic feed forward networks (SD-FFNN). The SD-FFNN is a hybrid model where training is based on a combination of simulated annealing with the powerful minima seeking conjugate gradient, which improves the likelihood of finding good extrema while containing enough determinism. Clearly, the problem can be approached from an evolutionary computation perspective with networks trained with evolution strategies, particle swarm optimization, and others. In particular genetic algorithms may be used.

The output of both the output layer and the last hidden layer are exported (Fig. 2) and are used for computing two different error measures.

One particular algorithm for multi-objective optimization is the elitist non-dominated sorting genetic algorithm (NSGA-II) \([4], [5], [6], [2]\). It has the features that it \( i \) uses elitism, \( ii \) uses an explicit diversity preserving mechanism, and \( iii \) emphasizes the non-dominated solutions.

### IV. Multi-objective Optimization of Neural Networks for Space Transformation

In the supervised case, a natural choice for the \( l \) mapping is an NDA neural network \([28], [16], [17], [13]\). The classical backpropagation approach to building NDA networks suffers from the well-known problem of local entrapment. This problem was approached in \([27]\) with hybrid stochastic-deterministic feed forward networks (SD-FFNN). The SD-FFNN is a hybrid model where training is based on a combination of simulated annealing with the powerful minima seeking conjugate gradient, which improves the likelihood of finding good extrema while containing enough determinism. Clearly, the problem can be approached from an evolutionary computation perspective with networks trained with evolution strategies, particle swarm optimization, and others. In particular genetic algorithms may be used.

The output of both the output layer and the last hidden layer are exported (Fig. 2) and are used for computing two different error measures.
layers of the network. In this paper, the Sammon Error (Eq.1) was the measure used for characterizing dissimilarity loss.

V. GENETIC PROGRAMMING

Analytic functions are among the most important building blocks for modeling, and are a classical form of knowledge. Direct discovery of general analytic functions can be approached from a computational intelligence perspective via evolutionary computation. Genetic programming techniques aim at evolving computer programs, which ultimately are functions. Among these techniques, gene expression programming (GEP) is appealing [9]. It is an evolutionary algorithm as it uses populations of individuals, selects them according to fitness, and introduces genetic variation using one or more genetic operators. GEP individuals are nonlinear entities of different sizes and shapes (expression trees) encoded as strings of fixed length. For the interplay of the GEP chromosomes and the expression trees (ET), GEP uses a translation system to transfer the chromosomes into expression trees and vice versa [9]. The set of operators applied to GEP chromosomes always produces valid ETs.

The chromosomes in GEP itself are composed of genes structurally organized into a head and a tail [8]. The head contains symbols that represent both functions (from a function set F) and terminals (from a terminal set T), whereas the tail contains only terminals. Two different alphabets occur at different regions within a gene. For each problem, the length of the head h is chosen, whereas the length of the tail t is a function of h and the number of arguments of the function with the largest arity.

As an example, consider a gene composed of the function set \( F = \{ \mathcal{Q}, +, -, *, / \} \), where \( \mathcal{Q} \) represents the square root function, and the terminal set \( T = \{ a, b \} \). Such a gene (the tail is shown in bold) is: *Q−b+a+/−baababaaab., and encodes the ET which corresponds to the mathematical equation 
\[
 f(a,b) = \sqrt{b} \cdot \left( (a + \frac{b}{2} ) - ((a - b) + b) \right) \] simplified as 
\[
 f(a,b) = b \sqrt{b} \frac{b}{a}
\] 

GEP chromosomes are usually composed of more than one gene of equal length. For each problem the number of genes as well as the length of the head has to be chosen. Each gene encodes a sub-ET and the sub-ETs interact with one another forming more complex multi-subunit ETs through a connection function. To evaluate GEP chromosomes, different fitness functions can be used.

VI. APPLICATION TO EARTH SCIENCES: GEOPHYSICAL PROSPECTING

The previously described approach was applied to geophysical data obtained from an investigation dealing with the detection of underground caves. Karstification is a peculiar geomorphological and hydrogeological phenomenon produced by rock solution as the dominant process. As a consequence, the earth’s surface is covered by irregular morphologies, like lapiaz, closed depressions(dolinas), sinks, potholes and underground caves. The hydrographic network is usually poorly developed, and rain waters infiltrate to form an underground drainage system. Sometimes the caves are opened to the surface, but typically they are buried, requiring the use of geophysical methods. Cave detection is a very important problem in civil and geological engineering.

The studied area contained an accessible cave and geophysical methods complemented with a topographic survey were used with the purpose of finding their relation with subsurface phenomena [22]. This is a problem with partially defined classes: the existence of a cave beneath a measurement station is either known for sure if made over the known cave, or unknown since there might be a buried cave beneath. Accordingly, only one class membership is defined.

The set of geophysical methods included 1) the spontaneous electric potential (SP) of the earth’s surface measured in the dry season, 2) the vertical component of the electromagnetic field in the Very Low Frequency (VLF) region of the spectrum, 3) the SP in the rainy season, 4) the gamma radioactive intensity and 2) the local topography. These four physical fields, along with the surface topography, were the five variables to be used in the study. In the area, a gentle variation in geological conditions for both the bedrock and the overburden was suspected by geologists. An isolation of the different geophysical field sources was necessary in order to focus the study on the contribution coming from underground targets, in an attempt to minimize the influence of both the larger geological structures and the local heterogeneities.

A. Data Preprocessing

The complexity of these measured geophysical fields in the area is illustrated by the distribution of the standard scores corresponding to the gamma ray intensity and the surface topography. While radioactivity is highly noisy, topography shows few features. Both fields are shown in Fig.3 (left and right). Since the geophysical fields are measured using different units, in order to neutralize the effect of the different scales introduced by the units of measurement, all values were transformed to standard scores (i.e. to variables with zero mean and unit variance).

Each geophysical field was assumed to be described by the following additive two-dimensional model composed of trend, signal, and random noise: 
\[
 f(x,y) = t(x,y) + s(x,y) + n(x,y),
\] 
where \( f \) is the physical field, \( t \) is the trend, \( s \) is the signal, and \( n \) is the random noise component, respectively.

To isolate an approximation of the signals produced by the underground target bodies, a linear trend term \( \hat{t}(x,y) = c_0 + c_1x + c_2y \) was fitted (by least squares) and subtracted from the original field. The residuals \( \hat{r}(x,y) = f(x,y) - \hat{t}(x,y) \) were then filtered by direct convolution with a low pass finite-extent impulse response two-dimensional filter to attenuate the random noise component. Such convolution is given by 
\[
 \hat{s}(x,y) = \sum_{k_{1}=-N}^{N} \sum_{k_{2}=-N}^{N} h(k_{1},k_{2}) \hat{r}(x-k_{1},y-k_{2})
\] 
where \( \hat{r}(x,y) \) is the residual, \( \hat{s}(x,y) \) is the signal approximation, and 
\[
 h(k_{1},k_{2}) \text{ is the low-pass zero-phase shift filter.}
\]

The collection of residual fields (as an approximation to the signals for all physical fields) was used for analysis. In total, 1225 points in a regular grid were measured for the five physical fields previously mentioned. As a last preprocessing
step, the data was clustered with a simple clustering method (the leader algorithm [12]). This algorithm operates with a dissimilarity or similarity measure and a preset threshold. A single pass is made through the data objects, assigning each object to the first cluster whose leader (i.e. representative) is close enough (or similar enough) to the current object w.r.t. the specified measure and threshold. If no such matching leader is found, then the algorithm will set the current object to be a new leader; forming a new cluster. In particular, for heterogeneous data involving mixtures of nominal and ratio variables, the Gower similarity measure [11] has proven to be suitable. For objects $i$ and $j$ the similarity is given by

$$S_{ij} = \frac{\sum_{k=1}^{p} s_{ijk}}{\sum_{k=1}^{p} w_{ijk}}$$

(2)

where the weight of the attribute ($w_{ijk}$) is set equal to 0 or 1 depending on whether the comparison is considered valid for attribute $k$. If $v_k(i), v_k(j)$ are the values of attribute $k$ for objects $i$ and $j$ respectively, an invalid comparison occurs when at least one of them is missing. In this situation, $w_{ijk}$ is set to 0. For quantitative attributes (like the ones of the datasets used in the paper), the scores $s_{ijk}$ are assigned as $s_{ijk} = 1 - |v_k(i) - v_k(j)|/R_k$ where $R_k$ is the range of attribute $k$. For nominal attributes $s_{ijk} = 1$ if $v_k(i) = v_k(j)$, and 0 otherwise.

The Gower’s similarity measure was used with a threshold value of 0.97. As a result, 648 leaders (cluster representatives) were found, corresponding to a subset of the original data. This smaller data set retains most of the original similarity structure because of the high threshold value.

VII. MAIN RESULTS

A series of multi- and single objective experiments were performed in order to study some of the properties of the data used within this study.

The experimental settings for the multi-objective experiments are shown in Table-I, which comprise a description of the data, the leader algorithm options, the evolutionary multi-objective optimization options, and the two objective function parameters, including the parameters used for non-linear discriminant analysis.

The 3 multi-objective experiments each generate approximately 10 distinct multi-criteria solutions, which lead to a total of approximately 30 distinct solutions for the multi-criteria problem. Fig-4 shows the solutions and the resulting Pareto Front for the two objective functions (neural network classification error and sammon error of the constructed space w.r.t. the original space dissimilarity matrix) that were obtained by the NSGA-II algorithm[4]. Three solutions were selected from Fig-4 that represent the two extremes and a compromise of these two objectives. These selections were then each visualized by constructing a 3-dimensional VR space from the hidden layer of the neural network solutions as shown in Fig-5. The leftmost representation in Fig-5 shows the best multi-objective Sammon error solution, with the property of preserving data structure. While the rightmost representation shows the best multi-objective classification error solution; a space in which objects should be maximally separated in terms of their class membership (cave or unknown). The middle representation demonstrates a multi-objective compromised solution that attempts to both preserve the original data structure and separate the objects as much as possible w.r.t. class membership.
The extremal solutions found by multi-objective optimization as reported in Fig-4 were then compared to single-objective counterparts. In particular, the best structure preservation as reported in Fig-4 were then compared to single-objective optimization multi-objective solution was compared with a solution obtained by a single-objective optimization algorithm. The resulting 3-dimensional Fletcher-Reeves VR space is orthogonalized w.r.t. variance via principal component analysis and then plotted in the 3 dimensional non-linear discriminant analysis feed forward neural network was orthogonalized w.r.t. variance via principal component analysis and then plotted in the 3 dimensional space. The other pole of the space can be seen to contain another cluster of more densely packed objects. This other cluster represents objects that can be most dramatically non-linearly separated in terms of class structure from the cave objects, indicating their "non-caveness".

The extremal solutions found by multi-objective optimization algorithm, in this case Fletcher-Reeves [20]. The resulting 3-dimensional Fletcher-Reeves VR space is shown as the representation on the right of Fig-6. It has a lower Sammon error of 0.0266 compared to the multi-objective solution of 0.0266. The extremal solutions found by multi-objective optimization as reported in Fig-4 were then compared to single-objective counterparts. The best classification preservation multi-objective solution was compared to a solution obtained by a single-objective optimization algorithm. The resulting 3-dimensional Fletcher-Reeves VR space is orthogonalized w.r.t. variance via principal component analysis and then plotted in the 3 dimensional space. The other pole of the space can be seen to contain another cluster of more densely packed objects. This other cluster represents objects that can be most dramatically non-linearly separated in terms of class structure from the cave objects, indicating their "non-caveness".

The output of the last hidden layer of the NDA network trained with genetic algorithms in single-objective mode (targeting a 3D new space) can be described by:

\[
X = \Phi_1(v_1, v_2, v_3, v_4, v_5)
\]
\[
Y = \Phi_2(v_1, v_2, v_3, v_4, v_5)
\]
\[
Z = \Phi_3(v_1, v_2, v_3, v_4, v_5)
\]

where \(\{v_1, v_2, v_3, v_4, v_5\}\) are the original variables (i.e. the transformed physical fields), \(X, Y, Z\) are the variables in the new space, and \(\Phi_1, \Phi_2, \Phi_3\) are the non-linear functions of the original variables defining the mapping performed by the NDA neural network. The distribution of the data in the transformed space reveals a clearly polarized structure, which is further refined by the principal component (PC) transformation which was applied to the \(X, Y, Z\) variables of the new space, to create yet another new space (the principal components space).

\[
PC_1 = \psi_1(\Phi_1, \Phi_2, \Phi_3) = \psi_1 \circ \Phi
\]
\[
PC_2 = \psi_2(\Phi_1, \Phi_2, \Phi_3) = \psi_2 \circ \Phi
\]
\[
PC_3 = \psi_3(\Phi_1, \Phi_2, \Phi_3) = \psi_3 \circ \Phi
\]

where \(\psi_1, \psi_2, \psi_3\) are the classical linear functions involved in the PC transformation. As a whole, the process is expressed by a transformation given by function compositions.
Fig. 5. Selected multi-objective optimization algorithm (NSGA-II) solutions. Left: best Sammon error solution (Sammon error: 0.0383, Classif error: 0.5725). Right: best Classification error solution (Sammon error: 0.0506, Classif error: 0.5401). Middle: Solution compromising both error measures (Sammon error: 0.0422, Classif error: 0.5556). Dark objects represent measuring stations over the known surveyed cave location. Light objects represent measuring stations over locations in which it is not known whether a cave exists underground. Geometry = spheres, Behavior = static.

Fig. 6. Left: best Sammon error solution obtained by the multi-objective optimization algorithm. (Sammon error: 0.0383) Right: Fletcher-Reeves single-objective optimization Sammon error solution. (Sammon error: 0.0266) Dark and light object representation explained in Fig-5. Geometry = spheres, Behavior = static.

Fig. 7. Left: best Classification error solution obtained by multi-objective optimization. Right: Best single objective classification error solution obtained by nonlinear discriminant analysis for 3 dimensions; and then orthogonalized via principal component analysis. The objects representing measurements over the cave are located at the extreme right; with the exception of one known measurement located closer to the middle of the VR space. Dark and light object representation explained in Fig-5. Geometry = spheres, Behavior = static.
$\psi_i \circ \phi$, where $i \in [1, 3]$. It is clear from Table-IV that the first component accounts for about 97% of the total variance, thus explaining the separation of the elements of the cave and unknown classes.

**TABLE IV**

**PRINCIPAL COMPONENTS RESULTS FOR THE NDA 3-D SPACE. FOR ALL MODES, THE FIRST COMPONENT ACCOUNTS ABOUT 97% OF THE TOTAL VARIANCE.**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Property</th>
<th>PC-1</th>
<th>PC-2</th>
<th>PC-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>&quot;&quot;</td>
<td>2.3894</td>
<td>0.0882</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>&quot;&quot;</td>
<td>0.966</td>
<td>0.029</td>
<td>0.004</td>
</tr>
<tr>
<td>Covariance</td>
<td>&quot;&quot;</td>
<td>1.3830</td>
<td>0.0539</td>
<td>0.0086</td>
</tr>
<tr>
<td></td>
<td>&quot;&quot;</td>
<td>0.967</td>
<td>0.028</td>
<td>0.005</td>
</tr>
</tbody>
</table>

The NDA transformed space (Fig-7 right) shows a remarkable polarization of the pattern vectors in two half spaces. The rightmost half space contains all of the vectors corresponding to the known class (cave) which on the other hand have the positive values of the first PC. Moreover, almost all of the vectors of the cave class reside at the very extreme of the distribution, thus exhibiting the largest positive values of the first PC. The leftmost half space is composed only of elements of the unknown class, with negative values of the first component. The very nature of the NDA network indicates that these vectors correspond to measurements made on sites much less likely to have a cave underneath, and there is a densely packed cluster of such objects exhibiting the largest negative values. There is a much less dense middle space in between the two extremes representing the elements whose properties do not allow a clear distinction, therefore conforming an undecided region. Clearly, the value of the first principal component is a measure of the degree of "cavehood" expected for the corresponding vector of the transformed physical fields, and accordingly, its spatial distribution should give an indication about where to expect the presence of other caves, not yet opened to the surface. The distribution of the nonlinear composition $\psi_1 \circ \phi$ over the studied area is shown in Fig-8(left). This first principal component represents roughly 96.7% of the variance of the best non-linear discriminant neural network solution (which provides 3 non-linear attributes from the original 5) as found by the single objective (classification error minimization) optimization algorithm. There is a clear central ridge of high values coinciding with the zone where the known cave is located, and a fading halo as the distance from the cave increases, until it becomes almost nonexistent at the East-West borders. Geologically, this result is consistent with the nature of the karstification process, and the very fuzzy nature of geological boundaries, as known from ore and other kind of underground deposits. As such, the $\psi_1 \circ \phi$ function can be used as a base for constructing a data-driven fuzzy membership function for the cavehood property. In addition, it is interesting to observe the presence of an almost circular feature at the central-left portion of Fig-8, exhibiting high positive $\psi_1 \circ \phi$ values, suggesting the potential presence of buried caves in that area. A borehole drilled at that location hit a buried cavity.

**A. Gene Expression Programming Results**

In order to obtain a more precise mathematical description of the $\psi_1 \circ \phi$ function, genetic programming in the variant of GEP was applied with the purpose of finding a compact analytical representation of the composite mapping. Several experiments were performed targeting root mean squared error (RMSE) and relative absolute error (RAE) minimization. The results are shown in Table V

**TABLE V**

**RESULTS OF GENETIC EXPRESSION PROGRAMMING EXPERIMENTS.**

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>R</th>
<th>RAE</th>
<th>Fitness</th>
<th>Iter.</th>
<th>Target Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.326</td>
<td>0.9820</td>
<td>0.1737</td>
<td>748.209</td>
<td>18545</td>
<td>RMSE</td>
</tr>
<tr>
<td>2</td>
<td>0.5969</td>
<td>0.9530</td>
<td>0.2485</td>
<td>800.968</td>
<td>19094</td>
<td>RAE</td>
</tr>
<tr>
<td>3</td>
<td>0.5969</td>
<td>0.9372</td>
<td>0.2928</td>
<td>773.532</td>
<td>20380</td>
<td>RAE</td>
</tr>
<tr>
<td>4</td>
<td>0.5427</td>
<td>0.9477</td>
<td>0.2648</td>
<td>647.808</td>
<td>19880</td>
<td>RMSE</td>
</tr>
<tr>
<td>5</td>
<td>0.4708</td>
<td>0.9677</td>
<td>0.2251</td>
<td>698.931</td>
<td>19971</td>
<td>RMSE</td>
</tr>
</tbody>
</table>

In particular, Model-1 has low RMSE, as well as the highest correlation, therefore representing a good approximation to the combined action of the NDA neural network, in functional composition with the PC orthogonal transformation. The analytic approximation to $\psi_1 \circ \phi$ function is given by the expression:

$$f(\psi_1 \circ \phi)(v_1, v_2, v_3, v_4, v_5) = \frac{v_2 (k_1 + v_2)}{k_2 - v_4} + \frac{v_0}{\cos(v_4 + v_1 + \sin(v_2)) + k_3} + (\sin(v_1 + v_4) + v_4 + v_1) \sin(k_4)$$

with the constants $k_1 = 2.304687, k_2 = -6.591217, k_3 = -2.893525$ and $k_4 = 3.815918$.

The spatial distribution of this function is shown in Fig-8 (right), which approximates remarkably the original distribution obtained by the compositional application of the NDA network and the PC orthogonal transformation. The features corresponding to the known cave and the circular anomaly (where the previously unknown cave was found) are completely retained by the analytic function $\psi_1 \circ \phi$, evidencing the effectiveness of the GEP approach.

**VIII. Conclusions**

The combination of several computational intelligence approaches such as NDA neural networks, multi-objective optimization using genetic algorithms proved to be very effective for constructing new feature spaces for visual data mining. In particular, spaces oriented to maximize structure preservation (using the last hidden layer output of the NDA network) and classification accuracy (using the network’s output layer) simultaneously can be constructed using the set of solutions lying along the Pareto front, even for real world problems with partially defined classes. Complex
properties like those obtained by successive application of different nonlinear and linear mappings can be approximated effectively by genetic programming techniques like GEP.

ACKNOWLEDGMENT

The authors would like to thank Robert Orchard from the Integrated Reasoning Group (National Research Council Canada, Institute for Information Technology) for his constructive criticism of the first draft of this paper.

REFERENCES


