

Nonlinear Mapping using Particle Swarm Optimisation in Security based Applications

Auralia I. Edwards and Andries P. Engelbrecht
Department of Computer Science
University of Pretoria
South Africa
aedwards@cs.up.ac.za and engel@cs.up.ac.za

Abstract—Nonlinear mapping is an approach of multidimensional scaling where a high-dimensional space is transformed into a lower-dimensional space such that the topological characteristics of the original high-dimensional space are preserved. This enables visualisation and feature extraction of datasets. Problems exist in conventional mapping methods in that they can be slow or illustrate poor performance. Particle Swarm Optimisation (PSO) is a global optimisation approach that can effectively and efficiently map vectors from one dimension to another. A low-dimensional representation of vectors is used for classification purposes applicable in a number of security applications, such as face and hand recognition. Multimedia is regularly transmitted across the network, and needs to be protected. Watermarking has been designed to be such a protection scheme. The weakness of existing watermark schemes is briefly discussed. The ability to estimate the lower-dimensional watermarking subspace allows an attacker to remove the watermark. Currently, a linear mapping method for watermarking is used. A PSO will be a suitable substitution to efficiently estimate the watermark subspace.

I. INTRODUCTION

Visualisation makes it possible to interpret structures, groups and outliers that exist within data. Reducing the dimensionality of data without losing the original characteristics will lead to more efficient extraction of knowledge and lessen space and time requirements.

Linear and nonlinear mapping [18] are two major mapping models. Linear models are computationally efficient [10]. Nonlinear models, unlike linear models, are able to unfold nonlinear data manifolds. Each of these models can be further divided into: iterative, simultaneous, local and global categories. Mapping methods, more specifically nonlinear mapping methods, are renowned for not always obtaining an acceptable lower-dimensional representation in an acceptable number of iterations [20]. This is mainly due to the fact that the existing procedures used by mapping methods, such as gradient descent, are susceptible to premature convergence on a local optimum [6].

The Particle Swarm Optimisation (PSO) technique, developed by Kennedy and Eberhart [14], has shown to be less susceptible to premature convergence, making it an ideal approach to perform nonlinear mapping.

Digital watermarks were introduced to prevent data from being manipulated before distribution across the network.

Sincere thanks for the necessary funds from the Centre of Excellence for making this research possible.

Watermark violation refers to the ability that an unauthorised user detects, removes or disrupts the watermark [13]. If the watermark is degraded in any way such that its watermarking bits have been altered, then it will be ineffective in its effort to secure the content. Principle Component Analysis (PCA) is a linear method that has been used to estimate the watermark subspace. Showing that a PSO can efficiently perform mapping would be a faster approach.

The remainder of the paper is organised as follows: Section II provides an overview of nonlinear mapping methods. Section III describes the PSO algorithm along with a modified PSO, the Guaranteed Convergence PSO (GCPSO). Section IV describes the algorithm used for the experiments, corresponding PSO parameter selection, and various datasets. Experimental results are presented and discussed in Section V. Sections VI and VII show the important aspects and the problems that arise once an accurate lower-dimensional representation of the high-dimensional dataset is obtained. Face and hand recognition, as well as watermark security, is briefly discussed. Section VIII summarises the major findings and concludes the paper with a brief look at future work resulting from this research.

II. REDUCING THE DIMENSION

A. Background

Among the linear models of dimensionality reduction are Principal Component Analysis (PCA) [12], Projection Pursuit and Multidimensional Scaling (MDS) [5]. PCA finds the subspace that best preserves the variance of the data, Projection Pursuit generalises the PCA and MDS finds the subspace that best preserves the interpoint distances.

Among the local nonlinear models are Generative Topographic Mapping (GTM) [4], Local Linear Embedding (LLE) [23], Curvilinear Component Analysis (CCA) [6], Component Distance Analysis (CDA) [19], Stochastic Neighbour Embedding (SNE) [27], Kruskal [16] and Sammon mapping [25].

Calculating the mapping between two spaces can be computed using either a supervised or an unsupervised technique. The above nonlinear methods are unsupervised, since no information concerning the output of the projected space is given.

B. Problem Definition

Basically, nearby points in the higher dimensional space need to be nearby points in the lower space. If, in the higher-dimensional space, the distance between two vectors is length x , then the distance between these vectors must be proportional to x in the lower dimension. Therefore, the mapping error between two hyperspaces needs to be close to 0.

Two datasets are used to solve the problem of mapping vectors from one space to another. The first dataset is m -dimensional and the other is d -dimensional where $m \gg d$. Each dataset consists of n vectors. Information concerning the first dataset is represented as:

	\mathbf{x}_1	\mathbf{x}_2	\cdots	\mathbf{x}_m
1	a_1	b_1	\cdots	c_1
2	a_2	b_2	\cdots	c_2
\vdots	\vdots	\vdots	\vdots	\vdots
n	a_n	b_n	\cdots	c_n

which represents an $n \times m$ matrix where m is the number of columns, a_1 , b_1 and c_1 are real valued numbers presented as an m -dimensional vector: $a_1\mathbf{x}_1 + b_1\mathbf{x}_2 \cdots + c_1\mathbf{x}_m$. The second dataset is represented as an $n \times d$ matrix, where d is the target lower dimension, and each row of the matrix is generated to represent the corresponding m -dimensional vector in the lower d -dimensional space.

A lower left triangular ‘distance’ matrix is calculated to represent a topological distance between these two matrices.

Euclidean distance is a common method to calculate the distance between two points. However, a more accurate estimate of two points positioned on a curve will be obtained by calculating the length of the path along the manifold, referred to as the Geodesic or Curvilinear distance. This is calculated by dividing the path along the manifold into z parts. The total length of the path is approximated by adding the length of each of these z parts. By increasing z , a more accurate approximation of the actual length of the curve will be obtained. Refer to figure 1.

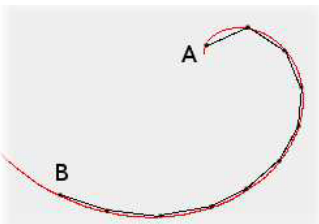


Fig. 1. Dividing the manifold into parts

One of the main problems concerning conventional nonlinear mapping methods is that a large number of experiments with different random initialisations have to be performed before the one with the lowest error is obtained [1]. As the number of vectors of the dataset increases, the computational requirement (time and storage) grows quadratically, referred to as the ‘curse of dimensionality’ [20] [3]. For example, Sammon’s mapping has a computational load of $O(n^2)$ [21][1]. The

minimum of Sammon’s mapping (error function) is achieved by carrying out a steepest-descent procedure.

A possible way to address the ‘curse of dimensionality’ problem is to use a Neural Network to perform a nonlinear mapping. Weight initialisations are usually done by using a computationally efficient technique such as a linear mapping method. A drawback to this approach is that a linear method cannot efficiently be used to train a dataset that is nonlinear to initialise weights [11].

A more robust optimisation algorithm that improves how a nonlinear mapping method finds an accurate lower-dimensional representation may provide a better solution to the ‘curse of dimensionality’ problem.

III. PARTICLE SWARM OPTIMISATION

PSO is a stochastic, population-based optimisation method which consists of a swarm of particles [14]. In nonlinear mapping, each particle represents a candidate solution in the lower $n \times d$ -dimensional search space [22].

All the solutions are evaluated by a fitness function that needs to be optimised. The particles ‘fly’ through the problem space by updating their velocities depending on its personal best position and the position encountered by the best particle in the neighbourhood thus far. The i^{th} particle’s new position is calculated using [28]:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t) \quad (1)$$

where $\mathbf{v}_i(t)$ represents the particle’s velocity and $\mathbf{x}_i(t)$ the current position. If d is the dimension index, the velocity is updated according to [28]:

$$v_{i,d}(t+1) = wv_{i,d}(t) + c_1r_{1,d}(t)(\mathbf{y}_{i,d}(t) - x_{i,d}(t)) + c_2r_{2,d}(t)(\hat{\mathbf{y}}_d(t) - x_{i,d}(t)) \quad (2)$$

where $0 \leq w < 1$ is an inertia weight determining how much of the particle’s previous velocity is preserved, c_1 and c_2 are acceleration constants, r_1 and r_2 are random variables sampled from $U(0, 1)$, \mathbf{y}_i is the personal best position found by the i^{th} particle and $\hat{\mathbf{y}}$ is the best position found by the entire swarm thus far (assuming *gbest* PSO). The second term models the particle’s personal knowledge about the search space, known as the cognitive component. The third term is known as the social component since information is obtained by interacting with other particles.

A. Neighbourhood Structures

Neighbourhood structures are used to determine the scope of information exchange between particles within the swarm.

The *gbest* PSO [8], based on the star topology, uses a fully connected network structure where each particle is attracted to the best solution found by the best particle within the entire swarm.

The *lbest* PSO [8], based on the ring topology, uses a one-dimensional lattice where all particles are aligned to have neighbours to the left and to the right.

The *Von Neumann* topology, developed by Kennedy and Mendes [15], is structured as a two-dimensional lattice such

that each particle can share information with particles not only to the left and right of it but also above and below it.

B. The Guaranteed Convergence PSO (GCP SO)

The GCP SO, introduced by Van den Bergh [28], addressed the problem of premature convergence on solutions that were not guaranteed to be local extrema. The velocity update equation for the global best particle (particle i) becomes

$$v_{i,d}(t+1) = wv_{i,d}(t) - x_{i,d}(t) + \hat{y}_d(t) + p(t)(1 - 2r_{2,d}(t)) \quad (3)$$

where r_d is a sequence of uniform random numbers sampled from $U(-1, 1)$ and $p(t)$ is a scaling factor. The p term causes the PSO to perform a search in an area surrounding the global best position $\hat{\mathbf{y}}$.

IV. EXPERIMENTAL PROCEDURE

The objective of this section is to show a PSO's capability to successfully map high-dimensional datasets to a lower dimension.

Section IV-A provides an overview of the algorithm used in the experiments. Section IV-B describes the error functions used to determine particle fitness. The datasets used in the experiments are discussed in IV-C.

A. Algorithm

Each particle in the swarm represents an $n \times d$ array in the solution space. The fitness of each particle is determined by a selected nonlinear mapping method's error function. The PSO algorithm aims to optimise the resulting fitness values until a suitable solution is found. All experiments were implemented using the Computational Intelligence Library (CILib) framework¹. An outline of the steps is given:

- 1) Each vector component within each particle is randomly initialised.
- 2) Evaluate the particle's fitness.
- 3) Update the particle's position to the best of either the current personal best or the current particle position.
- 4) Update both the particle's velocity and position according to equation (3) and equation (2) respectively.
- 5) Go to step 2 and repeat until a certain number of iterations has been exceeded.

The above algorithm is repeated 30 times. The average, standard deviation and a 95 % confidence interval of mapping error values were obtained.

B. Determining Fitness

The fitness function determines how well the representation in d -dimensional space approximates the topological structure of the given data in m -dimensional space. For the experiments, four different error functions are used:

- *Kruskal's Stress function:*

$$E_{Kruskal} = \sqrt{\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (\sigma_{ij} - d_{ij})^2}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sigma_{ij}^2}} \quad (4)$$

- *Sammon's Error function:*

$$E_{Sammon} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{(\sigma_{ij} - d_{ij})^2}{\sigma_{ij}}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sigma_{ij}} \quad (5)$$

- *Curvilinear Component Analysis (CCA):*

$$E_{CCA} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (d_{ij} - \sigma_{ij})^2 F(\sigma_{ij}) \quad (6)$$

- *Component Distance Analysis (CDA):*

CDA is a variant of CCA. The error function becomes:

$$E_{CDA} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\eta_{ij} - \sigma_{ij})^2 F(\sigma_{ij}) \quad (7)$$

where d_{ij} is the Euclidean distance between vector i and vector j in the m -dimensional space and σ_{ij} is the distance between vector i and vector j in the d -dimensional space. In equation (7), η_{ij} represents the curvilinear distance between vector i and vector j .

In equations (6) and (7), F is a monotonically decreasing function of σ_{ij} .

C. Datasets

Two datasets were used to investigate the performance of the PSO approach. The first dataset was chosen to analyse the influence when input dimension changes. The second dataset is more complicated and consisting of a significantly larger number of vectors.

1) *Dataset 1 - Multi:* The number of vectors in the dataset is represented by n , and m represents the dimension. Experiments were done with m taking the dimension values of 4, 6, 8 and 10. All of these are mapped to a 2-dimensional space that gives a straight line positioned at 45 degrees. Unless otherwise stated, the number of vectors (n) used throughout the experiments was 10.

TABLE I

DATASET 1 REPRESENTS THE HIGHER-DIMENSIONAL SPACE WHERE m IS THE DIMENSION AND n IS THE NUMBER OF VECTORS.

	1	...	m
0	0	...	0
1	1	...	1
⋮	⋮	⋮	⋮
n	n	...	n

2) *Dataset 2 - Swiss Roll:* The Swiss Roll dataset was generated (refer to figure 2) using:

$$r = \frac{3}{2}\pi(1 + 2 \times iter) \text{ where } iter = 0$$

$$x = r \times \cos(r); \quad (8)$$

$$y = y; \quad (9)$$

$$z = r \times \sin(r); \quad (10)$$

$$iter = iter + 1;$$

The points become closer towards the centre of the figure (refer to figure 2).

¹<http://cilib.sourceforge.net>

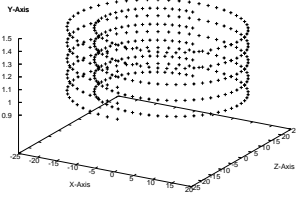


Fig. 2. Dataset 3: 500 vectors are used to display the 3-D Swiss Roll obtained by equations (8), (9) and (10)

D. Parameter Selection

The PSO parameter values, proposed by Van den Bergh [28], as listed in table II, have been used.

TABLE II
PSO PARAMETERS USED FOR THE EXPERIMENTS

Parameter	Value
$c_1 = c_2$	1.49618
inertia (w)	0.729844
$s_c = f_c$	5
neighbourhood size ($lbest$)	2

In addition to the parameters listed in table II, 40 particles and 5000 iterations were used for the experiments.

V. RESULTS AND DISCUSSION

The results obtained by the nonlinear mapping methods are investigated in Section V-A. The effect of changing the neighbourhood topologies is discussed in section V-B.

A. Nonlinear Mapping Methods

Table 4 summaries the results of Dataset 1, with an increasing input space dimension. Sammon's Mapping and CDA gave the best results. The most interesting aspect of these experiments is the fact that the error does not worsen as the dimension increased. In fact, in some cases the error becomes less. Figure 3 illustrates the best simulations of Dataset 1 (6 to 2 dimensional mapping) for each of the nonlinear mapping methods. Each method produced a straight line, with Kruskal's illustration being the least accurate.

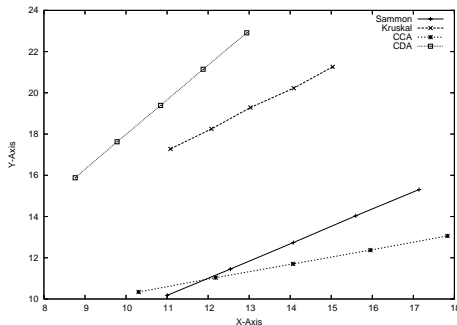


Fig. 3. Sammon, Kruskal, CDA and CCA illustrates dataset 1 (4 dimensional mapping to 2 dimensions). Only 5 points are shown.

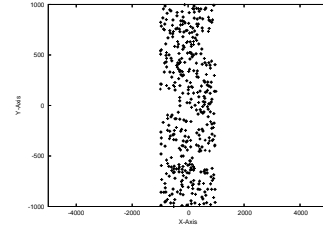


Fig. 4. Each point represents the coordinates in the lower dimension of the Swiss Roll dataset.

The results for the Swiss Roll dataset are provided in table IV. CDA gave the smallest confidence interval range.

Lee, et al. [18] used Isomap and CDA to project a Swiss Roll dataset of 20000 vectors to a 2-dimensional space, where the CDA method produced an S-shape. In the experiments, CDA produced an S-shape [18] (refer to figure 4), illustrating a correct mapping to the lower-dimensional space.

TABLE IV
DATASET 2 (SWISS ROLL) MAPPING FROM 3 TO 2 DIMENSIONS. 'A', 'B', 'C' AND 'D' REPRESENT KRUSKAL, SAMMON, CCA AND CDA RESPECTIVELY.

Method	Structure	Results
A	<i>gbest</i>	$9.81e^{-1} \pm 1.15e^{-4}$
	<i>Von Neumann</i>	$9.81e^{-1} \pm 2.25e^{-4}$
	<i>lbest</i>	$9.81e^{-1} \pm 2.03e^{-4}$
B	<i>gbest</i>	$9.53e^{-1} \pm 2.94e^{-4}$
	<i>Von Neumann</i>	$9.53e^{-1} \pm 4.18e^{-4}$
	<i>lbest</i>	$9.54e^{-1} \pm 2.89e^{-4}$
C	<i>gbest</i>	$4.30e^{-3} \pm 4.02e^{-3}$
	<i>von Neumann</i>	$3.26e^{-7} \pm 3.94e^{-7}$
	<i>lbest</i>	$1.22e^{-2} \pm 1.89e^{-2}$
D	<i>gbest</i>	$3.93e^{-2} \pm 4.68e^{-2}$
	<i>Von Neumann</i>	$1.16e^{-7} \quad 2.42 \pm 1.15e^{-7}$
	<i>lbest</i>	$4.92e^{-3} \pm 8.53e^{-3}$

B. Neighbourhood topologies

It was apparent that the *gbest* and *Von Neumann* topologies' average error decreased as the dimension of the dataset increased (refer to table III). The *gbest* topology converged quickly to a solution; better results occurred when the network structure was fully connected. A reason for this behaviour would be that the *gbest* simulations were less susceptible to local minima on datasets containing fewer vectors. However in table IV, the communication structures showed to have a larger influence when used with the CCA and CDA mapping methods. A larger dataset (Swiss Roll) that contained a more complex structure, illustrated *Von Neumann* to be the superior topology, since it produced the lowest errors and smallest confidence intervals.

VI. CLASSIFICATION APPLICATIONS

Mapping a high-dimensional image, such as a face, to a low-dimensional space before image classification is a common procedure. It is both time consuming and computationally

TABLE III

MAPPING DATASET 1 FROM INPUT DIMENSIONS OF 4, 6, 8 AND 10 TO A DIMENSION OF 2 IN THE OUTPUT SPACE. 'A', 'B', 'C' AND 'D' REPRESENT KRUSKAL, SAMMON, CCA AND CDA RESPECTIVELY.

	Structure	4 to 2	6 to 2	8 to 2	10 to 2
A	<i>gbest</i>	$1.39e^{-5} \pm 3.75e^{-6}$	$1.37e^{-5} \pm 3.92e^{-6}$	$1.12e^{-5} \pm 2.91e^{-6}$	$9.97e^{-6} \pm 3.00e^{-6}$
	<i>Von Neumann</i>	$3.39e^{-5} \pm 8.01e^{-6}$	$2.67e^{-5} \pm 1.06e^{-5}$	$2.71e^{-5} \pm 8.25e^{-6}$	$2.7e^{-5} \pm 8.27e^{-6}$
	<i>lbest</i>	$4.15e^{-5} \pm 1.30e^{-5}$	$5.01e^{-5} \pm 1.47e^{-5}$	$4.21e^{-5} \pm 1.59e^{-5}$	$5.30e^{-5} \pm 1.56e^{-5}$
B	<i>gbest</i>	$1.04e^{-11} \pm 6.32e^{-12}$	$2.41e^{-11} \pm 2.09e^{-11}$	$2.65e^{-11} \pm 1.17e^{-11}$	$1.22e^{-11} \pm 6.64e^{-12}$
	<i>Von Neumann</i>	$9.28e^{-11} \pm 5.38e^{-11}$	$2.07e^{-10} \pm 1.23e^{-10}$	$1.46e^{-10} \pm 9.05e^{-11}$	$1.97e^{-10} \pm 1.17e^{-10}$
	<i>lbest</i>	$6.54e^{-10} \pm 3.30e^{-10}$	$2.92e^{-10} \pm 2.09e^{-10}$	$3.74e^{-10} \pm 2.63e^{-10}$	$5.87e^{-10} \pm 4.92e^{-10}$
C	<i>gbest</i>	$1.41e^{-9} \pm 2.01e^{-9}$	$1.53e^{-9} \pm 1.59e^{-9}$	$1.29e^{-9} \pm 1.65e^{-9}$	$1.05e^{-9} \pm 9.44e^{-10}$
	<i>Von Neumann</i>	$5.06e^{-9} \pm 2.58e^{-9}$	$5.38e^{-9} \pm 5.64e^{-9}$	$5.92e^{-9} \pm 4.72e^{-9}$	$1.46e^{-8} \pm 1.63e^{-8}$
	<i>lbest</i>	$1.63e^{-8} \pm 7.26e^{-9}$	$1.91e^{-8} \pm 8.68e^{-9}$	$2.16e^{-8} \pm 1.34e^{-8}$	$1.68e^{-8} \pm 1.00e^{-8}$
D	<i>gbest</i>	$2.00e^{-10} \pm 7.96e^{-11}$	$5.80e^{-10} \pm 3.18e^{-10}$	$6.56e^{-10} \pm 3.59e^{-10}$	$9.79e^{-10} \pm 3.66e^{-10}$
	<i>Von Neumann</i>	$3.66e^{-9} \pm 2.65e^{-9}$	$3.73e^{-9} \pm 2.47e^{-9}$	$3.27e^{-9} \pm 1.57e^{-9}$	$5.47e^{-9} \pm 3.14e^{-9}$
	<i>lbest</i>	$8.86e^{-9} \pm 5.69e^{-9}$	$2.11e^{-8} \pm 1.13e^{-8}$	$1.33e^{-8} \pm 7.88e^{-9}$	$2.54e^{-8} \pm 2.25e^{-8}$

expensive to immediately categorise an image existing in its original high-dimensional space. This is especially the case when an image needs to be classified in real-time.

A PSO was successfully used to perform nonlinear mapping on a dataset that contains high dimensions (Dataset 1). A good lower-dimensional representation was obtained and notably in an acceptable time-frame.

A. Hands and Faces Applications

Today, fingerprint and retina recognition are used as effective security measures. PCA is a linear method and has been used in face recognition applications [17]. In [20], the CCA nonlinear mapping method produced better results than the PCA. CCA uses a gradient descent method to obtain its results, which as described earlier does not always guarantee acceptable results. It would be ideal to use a PSO in conjunction with the CCA's error mapping function to give the same, if not better results, such that hands or faces can be classified.

B. Fingerprint Applications

Matching in fingerprint analysis is usually based on lower-level features determined by singularities in the finger ridge pattern known as minutiae [24] [26]. A number of mapping techniques have been used to classify and analyse fingerprints. Lee's triangulation method has been used in fingerprint analysis. However, problems arise when it is required to project a large number of patterns in an efficient way. PSOs can be used to achieve such a mapping.

VII. WATERMARK VIOLATION

Watermarks are used to ensure authenticity of content. Unauthorized users should not be permitted to remove the embedded watermark. However, estimating the watermark's subspace will enable the user to remove it.

In [7], four different algorithms are described, namely Code Division Multiple Access (CDMA), Time Division Multiple Access (TDMA), Dither Index Modulation (DIM) and Embedding Strength Modulation (ESM). All of these algorithms can be reduced to a simple scheme expressed as [7]:

$$\mathbf{c}_w = \mathbf{c}_o + \sum_{i=1}^p \lambda_i w_i(K) \quad (11)$$

where c_o is the original digital content. K represents a secret key defining a private watermarking scheme which has fewer dimensions than the content that is to be protected. The random watermark patterns w_i 's are dependent on this key and the λ_i 's are dependent on their associating mixing coefficients [7]. To obtain back the hidden bits, the watermarked digital content is correlated with each one of the random watermark patterns using [7]:

$$p_i = \mathbf{c}_w \bullet \mathbf{w}_i(K) \quad (12)$$

where \bullet is a correlation operator. The embedded watermark is thus bounded to this private low-dimensional subspace.

A. Problem Definition

It has been shown, in previous work, that a watermark, when embedded in a digital document, can be estimated and remodulated [29]. The process in [7] describes the collision attack which can be divided into individual watermark estimation and watermark subspace estimation where the watermark can be removed.

1) *Individual watermark estimation*: Ideally, finding the watermark can be achieved by finding the difference between the watermarked content and its associated original one

$$\mathbf{w} = \mathbf{c}_w - \mathbf{c}_o \quad (13)$$

However, the original content is not known and the watermark will then need to be estimated [7]

$$\tilde{\mathbf{w}} = \mathbf{c}_w - \zeta(\mathbf{c}_w) \quad (14)$$

where ζ is a low-pass filter. Using equation (14) could lead to inaccurate estimators. However, using the above method it will be possible for a collection of n individual watermark estimates of size s to be obtained [7].

2) *Subspace Estimation*: Let $\tilde{\mathbf{W}}$ be a $s \times q$ matrix where the columns estimate the watermark \mathbf{w} . It would then be required to find a $p \times q$ matrix where $p \ll s$. p represents the dimension in lower space and s is the dimension in the higher space.

A nonlinear/linear estimate of the watermark subspace can be performed by a mapping technique. In [7], PCA was

chosen to perform the linear mapping. For more complicated watermarks, a nonlinear method would be more appropriate.

A set of vectors is obtained that span a subspace that is assumed to be close to the subspace of the watermark. Once this has been estimated, the watermark can be removed.

B. Using Particle Swarm Optimisation

Problems occur with standard approaches such as PCA or some of the other mapping methods in that they can be computationally complex and require a large amount of storage space. A PSO is less susceptible to these issues.

VIII. CONCLUSION AND FUTURE WORK

Four different nonlinear mapping methods were investigated and comparisons of these methods in spaces of various dimensionality was examined. The robust behaviour of the PSO technique was illustrated with errors being only marginally altered when dataset dimensions changed. PSOs can be used to perform nonlinear mapping.

Increasing the dimension had shown to be insignificant on datasets where the number of vectors were small.

The *Von Neumann* topology proved to be more superior on a more complex dataset (such as the Swiss Roll). It was apparent that particles needed to explore larger areas of the search space, without immediately being drawn to the global best particle, as with the *gbest* topology.

Further comparisons with other categories of nonlinear mapping methods will be analysed. These could include global approaches such as Isomap [2] and SSammons [21] (an extension to Sammon's Mapping that considers sparse data). Esquivel and Coello [9] developed a hybridisation of the standard PSO algorithm that has shown significant improvement over *lbest* and *gbest* topologies. Future work will include an investigation of this "non-parametric" PSO [8].

Applying the PSO to more complex datasets containing data such as facial features, hands or fingerprints will be investigated. Using the PSO as a substitution to the PCA for watermark estimation will also be experimentally analysed. The effectiveness will be determined within the authenticity verification of transmitted multimedia in various network applications.

REFERENCES

- [1] D. Agrafiotis and V. Lobanov, "Nonlinear Mapping Techniques," *Computer Science*, vol. 40, pp. 1356–1362, 11 April 2000.
- [2] M. Balasubramanian, E. Schwartz, and J. Tenenbaum, "The Isomap Algorithm and Topological Stability," *Science*, vol. 295, p. 7, January 2002.
- [3] R. Bellman, "Adaptive control processes: A guided tour," Princeton University Press, Princeton, Tech. Rep., 1961.
- [4] C. Bishop, G. Hinton, and I. Strachan, "GTM through time," *Fifth International Conference on Artificial Neural Networks*, pp. 111 – 116, July 1997.
- [5] T. Cox and M. Cox, "Multidimensional Scaling," Chapman and Hall, Tech. Rep., 1994.
- [6] Demartines and J. P. Herault, "Curvilinear Component Analysis: A Self-Organising Neural Network for Nonlinear Mapping of Data Sets," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148–154, January 1997.
- [7] G. Doerr and J. Dugelay, "Danger of low-dimensional watermarking subspaces," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 93–96, May 2004.
- [8] A. P. Engelbrecht, *Computational Intelligence: An Introduction*. Wiley and Sons, 2002, ch. 6, p. 191.
- [9] S. Esquivel and C. Coello, "On the use of particle swarm optimization with multimodal functions," *The 2003 Congress on Evolutionary Computation*, vol. 2, pp. 1130–1136, Dec 2003.
- [10] A. Hadid, O. Kouropteva, and M. Pietikainen, "Unsupervised Learning using Locally Linear Embedding: Experiments with Face Pose Analysis," *16th International Conference on Pattern Recognition*, vol. 1, pp. 111–114, 11 - 15 August 2002.
- [11] W. Hseih, "Nonlinear principal component analysis by neural networks," *Tellus*, vol. 53A, pp. 599–615, 2001.
- [12] I. Jolliffe, "Principal Component Analysis," 1986.
- [13] T. Kalker, "Considerations on watermark security," *Proceedings of the IEEE Workshop on Multimedia Signal Processing*, pp. 201–206, October 2001.
- [14] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *Proceedings of International Conference on Neural Networks*, pp. 1942–1948, 1995.
- [15] J. Kennedy and R. Mendes, "Population Structure and Particle Swarm performance," *In Proceedings of Congress of Evolutionary Computation, Hawaii USA*, 2002.
- [16] J. Kruskal, *Nonmetric multidimensional scaling: A numerical method*. Psychometrika, 1964, ch. 29, pp. 115–129.
- [17] S. Lawrence, C. Giles, A. Tsoi, and A. Back, "Face recognition: A convolutionary neural network approach," *IEEE Transactions on Neural Networks*, vol. 8, pp. 98–113, 1997.
- [18] J. Lee, A. Lendasse, and M. Verleysen, "Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis," *Neurocomputing*, vol. 57, pp. 49–76, 2004.
- [19] J. Lee, A. Lendasse, N. Donckers, and M. Verleysen, "A Robust Non-linear Projection Method," *European Symposium on Artificial Neural Networks Bruges (Belgium)*, vol. 8, pp. 13–20, April 2000.
- [20] R. Lotlikar and R. Kothari, "Face Recognition using Curvilinear Component Analysis," *World Congress on Computational Intelligence*, pp. 1778–1783, 1998.
- [21] M. Martin-Merino and A. Munoz, "A New Sammon Algorithm for Sparse Data Visualization," *17th International conference on Pattern Recognition*, vol. 1, pp. 477–481, 23–26 August 2004.
- [22] E. Peer, F. van den Bergh, and A. Engelbrecht, "Using Neighborhoods with the Guaranteed Convergence PSO," *Proceedings of the 2003 Swarm Intelligence Symposium*, pp. 235–242, 24–26 April 2003.
- [23] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [24] A. A. Salch and R. Adhami, "Curvature-based matching approach for Automatic Fingerprint Identification," *Southeastern Symposium on System Theory*, pp. 171 – 175, March 2001.
- [25] J. Sammon, "A Nonlinear Mapping algorithm for data structure analysis," *IEEE Trans. Computer*, vol. 18, no. 5, pp. 401–409, 1969.
- [26] W. Shalash and F. Abou-Chadi, "A fingerprint classification technique multilayer som," *Seventeenth National Radio Science Conference*, 2000.
- [27] B. Turhan, "Nonlinear Dimensionality Reduction methods for Pattern Recognition," Master's thesis, Graduate Program in Computer Engineering, Bogazici University, 2004.
- [28] F. van den Bergh, "An Analysis of Particle Swarm Optimizers," Ph.D. dissertation, University of Pretoria, Department of Computer Science, 2002.
- [29] C. Voloshynovskiy, S. Pereira, A. Herrigal, N. Baumgartner, and T. Pun, "Generalised watermarking attack based on watermark estimation and perceptual remodulation," *Security and Watermarking of Multimedia Contents*, vol. 3971, pp. 358–370, January 2000.

Auralia I. Edwards received the Hons-B.Sc degree, with distinction, in Computer Science in 2004 and a B.Sc degree, with distinction, in Computer Science in 2003. She started her M.Sc degree in Computer Science at the University of Pretoria in 2005.

Andries P. Engelbrecht received the B.Sc, Hons-B.Sc, M.Sc and PhD in Computer Science at the University of Stellenbosch in 1990, 1992, 1994, 1999 respectively. He is currently a full-professor in Computer Science at the University of Pretoria.