

Coventry University Repository for the Virtual Environment
(CURVE)

Author names: Jayne, C. , Lanitis, A. and Christodoulou, C.

Title: Jayne, C. , Lanitis, A. and Christodoulou, C.

Article & version: Post-print

Original citation & hyperlink:

Jayne, C. , Lanitis, A. and Christodoulou, C. (2012) One-to-many neural network mapping techniques for face image synthesis. *Expert Systems with Applications*, volume 39 (10)

<http://dx.doi.org/10.1016/j.eswa.2012.02.177>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Available in the CURVE Research Collection: July 2012

<http://curve.coventry.ac.uk/open>

One-to-many neural network mapping techniques for face image synthesis

C. Jayne

A. Lanitis

C. Christodoulou

Corresponding author

Department of Computing
Coventry University
Priory Street, Coventry CV1 5FB,
UK

Tel:+44(0)24 7688 8710

Fax:+44(0)24 7688 8030

Chrisina.Jayne2@coventry.ac.uk

Department of Multimedia and
Graphic Arts,
Faculty of Communication and
Applied Arts, Cyprus University
of Technology,
31 Archbishop Kyprianos
Street, P. O. Box 50329, 3603
Lemesos, Cyprus

andreas.lanitis@cut.ac.cy

Department of Computer
Science,
University of Cyprus,
75 Kallipoleos Avenue,
P.O. Box 20537, 1678
Nicosia, Cyprus

cchrist@cs.ucy.ac.cy

Abstract

This paper investigates the performance of neural network-based techniques applied to the problem of defining the relationship between a particular type of variation in face images and the multivariate data distributions of these images. In this respect the problem of defining a mapping associating a quantified facial attribute and the overall typical facial appearance is addressed. In particular the applicability of formulating a mapping function using neural network-based methods like Multilayer Perceptrons (MLPs), Radial Basis Functions (RBFs), Mixture Density Networks (MDNs) and a latent variable method, the General Topographic Mapping (GTM) is investigated. Quantitative and visual results obtained during the experimental investigation, suggest that for one-to-many problems, where the entire variance of the distribution is not required, the RBFs are the best options when compared to MLPs, MDNs and GTM. The proposed techniques can be applied to applications involving face image synthesis and other applications that require one-to-many mapping transformations.

Keywords: isolating sources of variation, neural networks, face image synthesis

Highlights:

- Neural network methods are applied for learning a mapping function that relates particular facial attributes to the overall facial appearance.
- A comparative evaluation of different neural network methods applied to the problem of one-to-many mapping is presented.
- Visual and quantitative results are demonstrated
- Application of one-to-many neural network methods to the problem of face image synthesis

1. Introduction

In many problems involving the analysis of multivariate data distributions, it is desirable to isolate specific sources of variation within the distribution, where the sources of variation in question represent a quantity of interest related to the specific problem domain. The isolation of different types of variation within a training set enables the generation of synthetic samples of the distribution given the numerical value of a single type of variation (or data dimension). Usually multiple parameters are required to specify a complete sample in a distribution, thus the process of generating a sample given the value of a simple parameter takes the form of one-to-many mapping. In general one-to-many mapping problems are ill-conditioned, requiring the use of dedicated techniques that use prior knowledge in attempting to formulate an optimized mapping function.

With our work we aim to investigate the use of different neural network methods for defining a mapping associating a specific source of variation within a distribution and a given representation of this data distribution. In particular we address the problem of defining a mapping associating a quantified facial attribute and the overall typical facial appearance, enabling in this way the synthesis of faces displaying certain facial attributes. Apart from the aforementioned application the findings of this work can be used in various tasks related to multivariate distributions including:

- Investigating the relationship between different types of variability within a training set and samples belonging to the same distribution.
- Categorization of the effects of hidden parameters.
- Exploration and analysis of multivariate face image distributions.

In practice, a typical multivariate data distribution is often represented by a finite set of samples characterised through the values of a set of parameters $\{\mathbf{b}^i = (b^i_1, \dots, b^i_d)\}_{i=1, N}$, where $\mathbf{b}^i \in Y$, a subset of \mathbb{R}^d . A source of variation is characterised through a finite set of numbers $\{q^i\}_{i=1, N}$, where $q^i \in X$, a subset of \mathbb{R} . These samples form a training set of corresponding input-output pairs $\{q^i, \mathbf{b}^i\}_{i=1, N}$. We consider the formulation of a mapping function that associates the inputs $\{q^i\}_{i=1, N}$, with the outputs $\{\mathbf{b}^i\}_{i=1, N}$ so that in some optimal sense the problem becomes a problem of mapping approximation (Carreira-Perpiñán, 2001). In this mapping, the dimension of the target data space is higher than the dimension of the input space and the target data could be multi-valued, i.e., for

some input the target output may not be unique. Although this leads to solving of a so-called mapping inversion problem (Carreira-Perpiñán, 2001), the aim is not to find a complete description of the data for the purpose of predicting the outputs corresponding to new input vectors, but to find a mapping that generates typical samples of the distribution given specific input values of the quantity of interest.

In the context of the problem of reconstructing face images displaying certain facial attributes, Principal Component Analysis (PCA)-based coding techniques (Edwards et al., 1998) are used to represent face images as vectors in a low dimensional space. Based on this representation, we examine the efficiency of methods that learn the mapping between certain quantified facial attributes and the coded face images in the training set. The resulting mapping function can be used for generating face images displaying faces with specific attributes. This application can be used as the basis for implementing systems that allow the synthesis of faces displaying specific facial attributes to a pre-defined extent. As part of this study, the facial attributes considered include the aspect ratio of a face, the distance between eyes, mouth width, nose length and age.

The problem of isolating specific sources of variation of face images has been considered by a numbers of researchers in previous studies. In (Lanitis, Taylor, Cootes, 1997), the main types of shape and texture variation exhibited in a set of face images are extracted by using Principal Component Analysis. In this context eigen-vectors associated with the highest variance within a training set are linked to specific types of facial deformations. However, in such cases it is not guaranteed that there will be one-to-one correspondence between eigen-vectors and distinct types of variations. In other occasions (Lanitis, 2003) regression is used for establishing dedicated functions that link facial appearance to specific facial attributes, allowing in that way the generation of faces displaying specific facial attributes. However, regression is suited only in the cases that attributes can get continuous values within a range of possible values. In the case that facial attributes are associated with discrete values belonging to specific classes (for example id and gender related attributes), the use of discriminant analysis has been investigated (Belhumeur, Hespanha & Kriegsmann, 1997). However, discriminant analysis-based methods have mainly been used for implementing face image classification rather than face image synthesis related applications.

In this paper, the use of network-based methods that can be used for isolating specific sources of variation within a training set and the subsequent generation of faces displaying specific attributes

is investigated. In particular, the following neural network-based methods are considered: Multilayer Perceptron (MLP) (Rumelhart, Hinton & Williams, 1986), Radial Basis Functions (Powell, 1985), Mixture Density Networks (MDN) (Bishop, 1994, Bishop, 1995) and the non-linear latent variable method Generative Topographic Mapping (GTM) (Bishop, Svensén & Williams, 1998). As a reference benchmark of the prediction accuracy we consider the values of the predicted variables that correspond to the average values over certain intervals of the quantity of interest that we are trying to isolate (the so called Sample Average (SA) method).

In many applications related to pattern classification or recognition, the tasks for feature extraction and dimensionality reduction are only intermediate steps towards solving the entire classification or recognition problems and these tasks attracted considerable research interest. However, the problem of isolating a source of variation from multivariate data distributions, and subsequently analysing the variation of this specific source within the distributions has not been extensively investigated in the research literature. We aim to address specifically this problem. The work presented in this paper builds on our previous work in the area where we attempted to use neural network one-to-many mapping methods in the problems of predicting individual stock share prices given the value of the general index and predicting the grades received by high school pupils, given the grade for a single course (Jayne, Lanitis & Christodoulou, 2011).

The rest of the paper is organised as follows: section 2 gives an overview of the relevant literature; section 3 illustrates brief theoretical background of the investigated neural network methods; section 4 describes the face synthesis application; section 5 presents the experiments, visual and quantitative results and section 6 provides the conclusions.

2. Literature Review

There exist well-established neural network methods for solving the mapping approximation problem such as the Multilayer Perceptron (MLP) (Rumelhart, Hinton & Williams, 1986) and Radial Basis Functions (RBF) (Powell, 1985). The aim of the training in these methods is to minimize a sum-of-square error function so that the outputs produced by the trained networks approximate the average of the target data, conditioned on the input vector (Bishop, 1995). It is reported in (Bishop, 1994) and (Richmond, 2001), that these conditional averages may not provide a complete description of the target variables especially for problems in which the mapping to be learned is multi-valued and the aim is to model the conditional probability distributions of the target variables (Bishop, 1994). In this paper, despite the fact that we have a

multi-valued mapping, we aim to model the conditional averages of the target data, conditioned on the input that represents a source of variation within this distribution. The idea is that when we change the value of the parameter representing the source of variation in the allowed range, the mapping that is defined will give typical representation of the target parameters exhibiting the isolated source of variation.

Bishop, 1994 introduces a new class of neural network models called Mixture Density Networks (MDN), which combine a conventional neural network with a mixture density model. The mixture density networks can represent in theory an arbitrary conditional probability distribution, which provides a complete description of target data conditioned on the input vector and may be used to predict the outputs corresponding to new input vectors. Practical applications of feed forward MLP and MDN to the acoustic-to-articulatory mapping inversion problem are considered in (Richmond, 2011). In this paper, it is reported that the performance of the feed-forward MLP is comparable with results of other inversion methods, but that it is limited to modelling points approximating a unimodal Gaussian. In addition, according to Richmond, 2001 the MLP does not give an indication of the variance of the distribution of the target points around the conditional average. In the problems considered in (Bishop, 1994) and (Richmond, 2001) the modality of the distribution of the target data is known in advance and this is used in selecting the number of the mixture components of the MDN.

Other methods that deal with the problem of mapping inversion and in particular mapping of a space with a smaller dimension to a target space with a higher dimension are based on latent variable models (Bartholomew, 1987). Latent variables refer to variables that are not directly observed or measured but can be inferred using a mathematical model and the available data from observations. Latent variables are also known as hidden variables or model parameters. The goal of a latent variable model is to find a representation for the distribution of the data in the higher dimensional data space in terms of a number of latent variables forming a smaller dimensional latent variable space. An example of a latent variable model is the well-known factor analysis, which is based on a linear transformation between the latent space and the data space (Bishop, 1995). The Generative Topographic Mapping (GTM) (Bishop, Svensén & Williams, 1998) is a non-linear latent variable method using a feed-forward neural network for the mapping of the points in the latent space into the corresponding points in the data space and the parameters of the model are determined using the Expectation-Maximization (EM) algorithm (Dempster, Larid & Rubin, 1977). The practical implementation of the GTM has two potential problems: the

dimension of the latent space has to be fixed in advance and the computational cost grows exponentially with the dimension of the latent space (Carreira-Perpiñán, 1999).

Density networks (MacKay & Gibbs, 1998) are probabilistic models similar to GTM. The relationship between the latent inputs and the observable data is implemented using a multilayer perceptron and trained by Monte Carlo methods. The density networks have been applied to the problem of modelling a protein family (MacKay & Gibbs, 1998). The biggest disadvantage of the density networks is the use of the computer-intensive sampling Monte Carlo methods, which do not scale well when the dimensionality is increased.

Even though the problem we consider in this paper bears similarities with the problem of sensitivity analysis with respect to neural networks, there are also distinct differences. In sensitivity analysis the significance of a single input feature to the output of a trained neural network is studied by applying that input, while keeping the rest of the inputs fixed and observing how sensitive the output is to that input feature (see for example Zeng & Yeung, 2003 and references therein). In the problem investigated in this paper, we do not have a trained neural network, but a model with parameterised representation of a face image. Based on our knowledge of the application, i.e., the different face parameters, we isolate a specific source of variation and carry out an one-to-many mapping between that isolated source and the face model (which is a multivariate data distribution). This allows the analysis of the variation of the isolated source within the face model.

3. Methods

In this section the basic theoretical background of each of the methods under investigation is presented.

3.1 Multilayer Perceptron Method (MLP)

The Multilayer Perceptron (MLP) (Rumelhart, Hinton & Williams, 1986) is the most widely used neural network architecture. Typically, it consists of three layers of neurons: input, hidden and output layers fully connected with adaptive weights. The input layer passes the input values through the hidden layer while the hidden and output layer neurons are active, i.e., they contain an activation function. The activation function of the hidden layer neurons is a smooth nonlinear function (e.g., sigmoid or hyperbolic tangent). The activation function in the output layer for

regression problems is usually a linear function (Bishop, 1995). During training, the input patterns are propagated through the network and the weights are adjusted to minimise the sum-of-square error function using a gradient descent process. In this paper a MLP with the scaled conjugate gradient algorithm is used (Møler, 1993), which combines the model-trusted approach and the conjugate gradient approach. It uses a numeric approximation for the second derivatives (Hessian matrix) to reduce the computations in the line search used by the traditional conjugate gradient algorithm.

3.2 Radial Basis Functions (RBF)

The Radial Basis Functions (RBF) (Powell, 1985) architecture is similar to the MLP architecture consisting of input, hidden and output layers of neurons, but there are some differences. Firstly the outputs in the hidden layer are not the product of the input pattern vector and the weight vector. Each neuron in the hidden layer is a centre of a cluster in the input data space found using a clustering algorithm (e.g., k-means, Duda & Hart, 1973). Secondly the transfer function associated with each hidden neuron is known as a radial basis function typically a Gaussian curve, through which the Euclidean distance between the input vector and the centre vector is passed. The output layer uses linear activation function and the weights between the hidden and output layer as well as the position of the centres and standard deviations of the Gaussian activation functions of the hidden layer neurons are adjusted using a gradient descent algorithm. The latter two parameters could also be optimised through the use of other optimisation algorithms.

3.3 Mixture Density Networks (MDN)

The Mixture Density Network (MDN) (Bishop, 1994, 1995) consists of a feed-forward neural network which maps the input vector to the control parameters of a mixture model of Gaussian components. The mixture model represents the conditional probability density function of the target variables, conditioned to the input vector of the neural network. The density function is a sum of a specified number of mixture components, each of which is the product of a Gaussian kernel function and a mixing coefficient. The mixing coefficients can be considered as prior probabilities of the target vector having been generated by the i^{th} kernel. Any neural network with universal approximation capabilities (e.g., MLP) can be used to map the input vector to the mixture model parameters.

3.4 Generative Topographic Mapping (GTM)

The Generative Topographic Mapping (GTM) (Bishop, Svensén & Williams, 1998) is a probability density model based on a constrained mixture of Gaussians whose parameters are typically optimised using the EM algorithm. The model describes the distribution of data in a space of several dimensions in terms of a smaller number of latent variables. The latent space is mapped to a non-Euclidean manifold in the data space by a non-linear function. This mapping can be constructed using a feed-forward neural network (e.g., RBF). Typically the GTM model is applied to data visualization, by inverting the transformation from latent space to data space using the Bayes' theorem. However, in this paper the GTM is applied only for finding the mapping from one dimensional latent space to the larger dimensional target data space.

4. Isolating sources of facial variation

We attempt to isolate sources of facial variation within a training set so that it is possible to generate face images displaying a certain facial attribute. In this context the mapping between the values of facial attributes and a low dimensional coded representation of faces is learnt, so that it is possible to synthesize a face once the numerical value of an attribute is fixed. For the experiments related to face reconstruction described in this paper we have used the FG-NET Aging Database (Lanitis, 2008). The FG-NET Aging Database is a publicly available image database containing face images showing a number of subjects at different ages. The database contains 1002 images from 82 different subjects with ages ranging between newborns to 69 years old subjects. For each image in the database a data file containing the locations of 68 facial landmarks (see figure 1) is available. Further information for each face image, including the actual age of the person shown in the image, is also available.

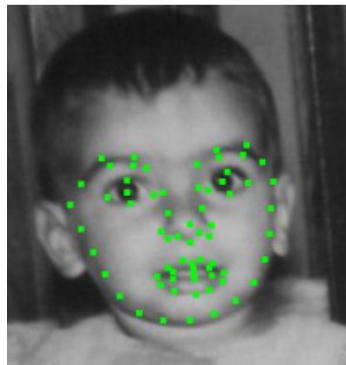


Figure 1: Location of the landmarks (68 points) on a typical face

Cootes, Edwards & Taylor, 2001 and Edwards et al., 1998 describe the generation of a statistical face model that can be used as a basis for coding face images in a low dimensional space. The process of generating such models involves the application of PCA on shape-normalized facial textures and coordinates of a number of landmarks located on each face (see figure 1). The main advantage of this type of models is their ability to map face images into a low dimensional space so that the process of manipulating face images takes the form of processing vectors containing model-based face representations. Usually each parameter in the low dimensional space explains different types of appearance variation encountered in the training set, such as expression, rotation, and inter-individual facial variation (Lanitis, Taylor & Cootes, 1997).

We have generated a face model using 1002 face images of 82 individuals and as a result we have a reversible compact coding of each face image as a set of 58 model parameters (see Figure 2).

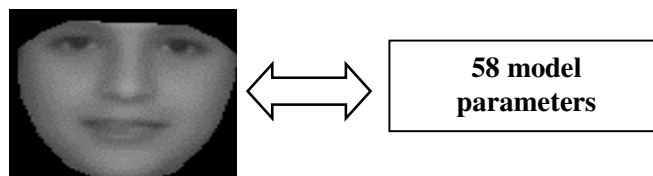


Figure 2: Reversible, compact coding of each face image using 58 model parameters

Based on the representation described above, we apply methods that learn the relationship between certain facial attributes and the coded face images in the training set. We then use the resulting mapping functions to reconstruct face images displaying faces with specific attributes. Figure 3 illustrates the set of attributes that we consider and the sequence of steps for reconstructing a face given a particular quantity representing the attribute in question.

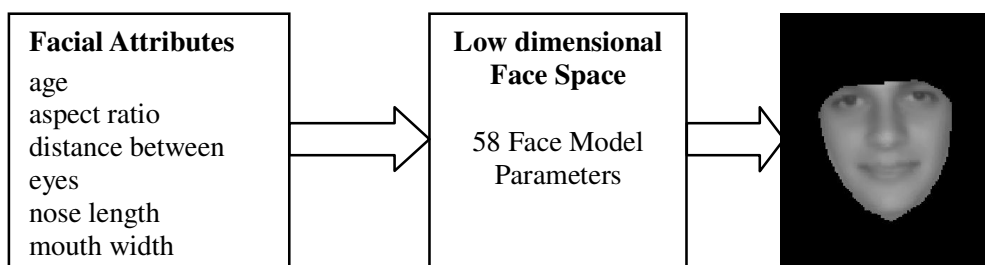


Figure 3: Sequence of steps for reconstructing a face given a quantified attribute. Once a numerical value of a face attribute is defined we use a one-to-many mapping function to estimate the values of the 58 model parameters, which are subsequently used for reconstructing a face instance.

For each facial attribute a unique mapping function is defined. In our experiments we consider the following attributes: age, aspect ratio, distance between eyes, mouth width and nose length. The age of each subject shown in our images was already available. For each face in the FG-NET dataset it is possible to estimate the actual values of the quantified facial attributes considered in our experiments (aspect ratio, distance between eyes, mouth width and nose length) by making use of the coordinates of the landmarks (see Figure 1) located on the training images. Figure 4 illustrates the landmark points used as the basis for estimating the values of different attributes and Table 1 provides details related to the quantification of different facial attributes, based on the locations of the points shown in Figure 4.

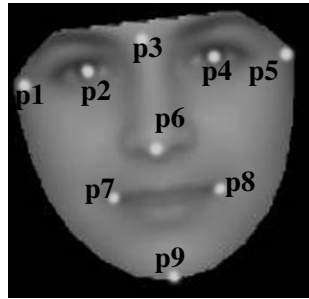


Figure 4: Points used for calculating the facial features

Table 1: Calculating face features

Feature	Description
Age	The age is quoted during collection of the images
Aspect ratio	The aspect ratio is calculated as the width (distance between points p1 and p5) of the face over the height (distance between points p3 and p9) of the face.
Distance between eyes	The distance between eyes is calculated by considering the distance between the two eyes (points p2 and p4) over the width of the face at the height of the eyes.
Nose length	The length of the nose is the distance between the tip of the nose (point p6) and the point between the eyebrows (point p3) on the level of the eyes over the height of the face.
Mouth width	The width of the mouth is calculated as the width of the mouth (distance between points p7 and p8) over the width of the face.

In effect the aim of our work in this case is to define a complex mapping function (g) that relates the numerical values of the specific facial attribute and the face model parameters, so that given a numerical value of facial attribute a typical face displaying that attribute can be reconstructed.

$$\mathbf{b} = g(q) \quad (1)$$

In this application the mapping function is defined by training MLP, RBF and MDN models using the numerical values of each attribute at the input and the 58 face model parameters at the output. It is worth noting that based on the work described in (Edwards et al., 1998) it is possible to use the values of the 58 parameters in conjunction with the related eigenvectors in order to generate a complete facial representation that includes both facial texture and the location of the 68 landmark points. The Sample Average (SA) method is applied by calculating the average vectors of the 58 model parameters \mathbf{b} corresponding to the numerical values of the facial attributes in selected intervals. We have also carried out an experiment using the Generative Topographic Mapping (GTM) approach for isolating one latent variable.

5. Experiments, Results and Discussion

5.1 Experimental methodology

For the experiments we use 498 images (belonging to subjects with ids 001 - 040) from the FG-NET Aging Database (Lanitis, 2008) for training and the remaining 504 images (belonging to subjects with ids 041 - 082) for testing. First we train neural network models using the MLP with the scaled conjugate gradient algorithm (Møler, 1993), the RBF and MDN methods. The inputs for the neural network model are the numerical values of a quantity q of interest and the outputs are the vectors of model parameters corresponding to faces images from the training set. For example, in our experiments q corresponds to a facial attribute such as age, aspect ratio, distance between eyes, mouth width and nose length and the output corresponds to the 58 model parameters of the corresponding face. In the MLP model, the network has one input node, one hidden layer with hyperbolic tangent (\tanh) activation function and an output layer with linear activation function, since the problem we consider is a regression problem. In the case of RBF similarly to the MLP, the input layer has one node, the output layer has linear outputs and the hidden layer consists of nodes (centres) with Gaussian basis functions. The Gaussian basis function centres and their widths are optimised by treating the basis functions as a mixture model and using the EM algorithm for finding these parameters. The number of hidden nodes in the MLP and RBF networks and the learning rate in the MLP network are set empirically. We also set empirically the number of hidden nodes and kernel functions (mixture components) in the MDN

model. Theoretically by choosing a mixture model with a sufficient number of kernel functions and a neural network with a sufficient number of hidden units, the MDN can approximate as closely as desired any conditional density. In the case of discrete multi-valued mappings the number of kernel functions should be at least equal to the maximum number of branches of the mapping. We have 58 outputs and for each input value we have a number of face images corresponding to that input. The dimensionality of the problem becomes very large if we select more than 2 kernel functions.

We have also carried out an experiment for isolating one latent variable using the GTM (Bishop, Svensén & Williams, 1998). The GTM models consist of an RBF non-linear mapping of the latent space density to a mixture of Gaussians in the data space of parameters (58 model parameters). The models are trained using EM algorithm. After training we use the RBF mapping to obtain the parameters corresponding to several values of the latent variable. We reconstruct the face images corresponding to those model parameters which show the variation that the isolated latent variable exhibits in the data space.

The SA method is applied by calculating the average vectors of the output parameters corresponding to the numerical values of the quantity q in selected intervals. For example, we calculate the average vectors of the 58 model parameters \mathbf{b} in seven equal sub-intervals between the minimum and the maximum value of each quantity q (i.e. for each facial attribute considered in our experiments).

5.2 Results and Discussion

After the training of the different models is completed, we vary the quantities for each attribute between its minimum and maximum values and at each step we reconstruct the corresponding face image. To evaluate the potential of the proposed approach in quantitative terms we calculate the values of each pre-specified attribute on reconstructed face images and compare it with the values of the attribute used as input to the trained networks. Tables 2-5 present the quantitative results showing the variation of the quantities for aspect ratio, distance between eyes, mouth width and nose length on the face images reconstructed from the obtained set of parameters, using the MLP, RBF, MDN and Sample Average (SA) methods.

Table 2: Quantitative results – aspect ratio

Aspect ratio	Absolute value errors			
	MLP	MDN	RBF	SA
0.6213	0.031957	0.296974	0.000802	0.001572
0.7112	0.001065	0.207073	0.000788	0.016278
0.8012	0.000546	0.117072	0.001428	0.009833
0.8911	0.001543	0.02717	0.000535	0.004772
0.9811	0.000157	0.06283	0.001509	0.009678
1.071	0.005977	0.152731	0.002587	0.019269
1.1609	0.038384	0.242632	0.002063	0.017953
Mean Error	<i>0.011376</i>	<i>0.158069</i>	<i>0.001388</i>	<i>0.011337</i>

Table 3: Quantitative results – distance between eyes

Distance between eyes values	Absolute value errors			
	MLP	MDN	RBF	SA
0.3711	0.003093	0.079969	0.000512	0.006402
0.3989	0.000116	0.052169	0.002584	0.006046
0.4268	0.000652	0.024268	0.002231	0.004916
0.4547	0.001293	0.003632	0.001062	0.001065
0.4826	0.002384	0.031532	0.001415	0.005533
0.5105	0.011617	0.059432	0.005278	0.007978
0.5384	0.025334	0.087332	0.009768	0.014736
Mean Error	<i>0.006356</i>	<i>0.048333</i>	<i>0.003264</i>	<i>0.006668</i>

Table 4: Quantitative results – mouth width

Mouth width values	Absolute value errors			
	MLP	MDN	RBF	SA
0.2253	0.00457	0.142519	0.003755	0.019947
0.2849	0.001034	0.082919	0.000944	0.001908
0.3444	0.000427	0.023418	0.000217	0.001658
0.404	0.000733	0.036182	0.000101	0.005298
0.4636	0.003323	0.095782	0.001057	0.009212
0.5232	0.01694	0.155382	0.000152	0.018931
0.5828	0.046321	0.214982	0.005881	0.005812
Mean Error	<i>0.010478</i>	<i>0.107312</i>	<i>0.00173</i>	<i>0.008966</i>

Table 5: Quantitative results – nose length

Nose length values	Absolute value errors			
	MLP	MDN	RBF	SA
0.239	0.004624	0.098894	0.005027	0.0122
0.275	0.002451	0.062894	0.000322	0.003895
0.3109	0.000436	0.026994	0.000593	0.002824
0.3469	0.001442	0.009006	0.001508	0.002122
0.3829	0.002281	0.045006	0.002435	0.004099
0.4189	0.002824	0.081006	0.002846	0.005005
0.4548	0.00332	0.116906	0.002028	0.003115
Mean Error	0.002483	0.062958	0.002108	0.004752

According to the quantitative results in all cases the RBF method exhibits the best results since when RBF is used the error between the input attribute value and the actual value of the attribute in the synthesized image is minimized. Statistical tests show that the mean errors obtained with the RBF for the different attributes are significantly different at the following levels: for aspect ratio 99.8%, for distance between eyes 88.7%, for mouth width 98.4% and for nose length 91.2%.

Figures 5-8 present the face images reconstructed from the obtained sets of parameters by varying the values of the aspect ratio, distance between eyes, mouth width and nose length using the methods under evaluation. The images shown in Figures 5-8 show clearly a trend of increasing the attribute in question in successive images.

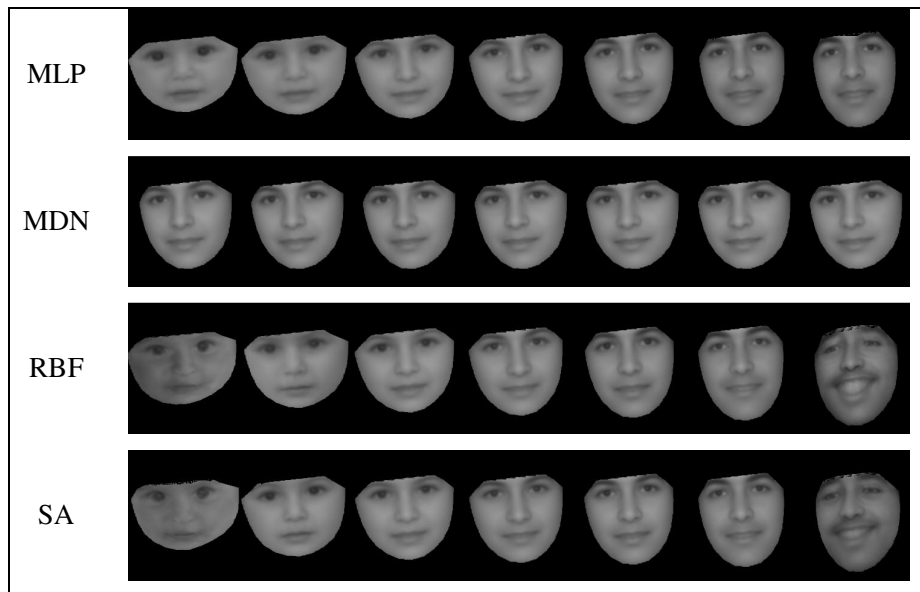


Figure 5: Variation of aspect ratio using MLP, MDN, RBF networks and SA method

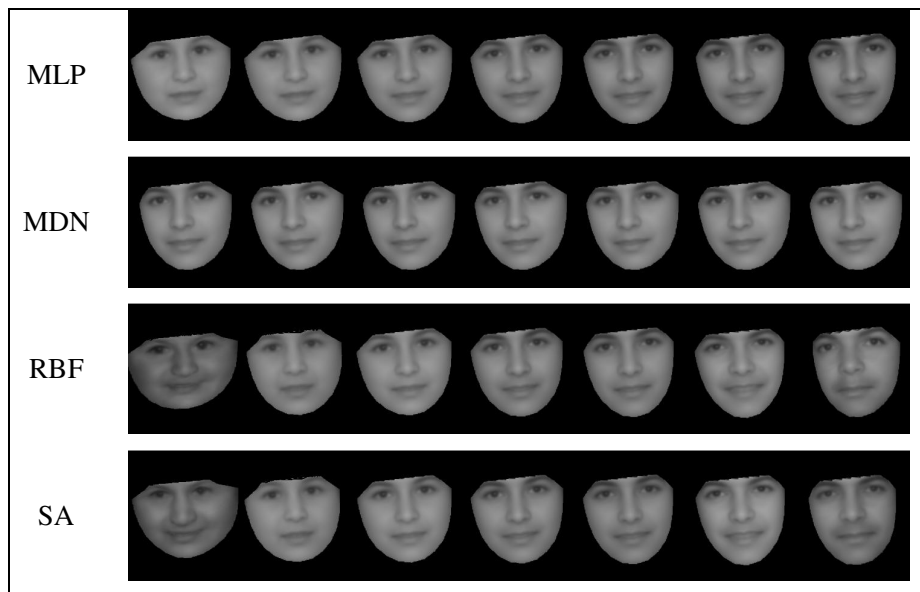


Figure 6: Variation of distance between eyes using MLP, MDN, RBF networks and SA method

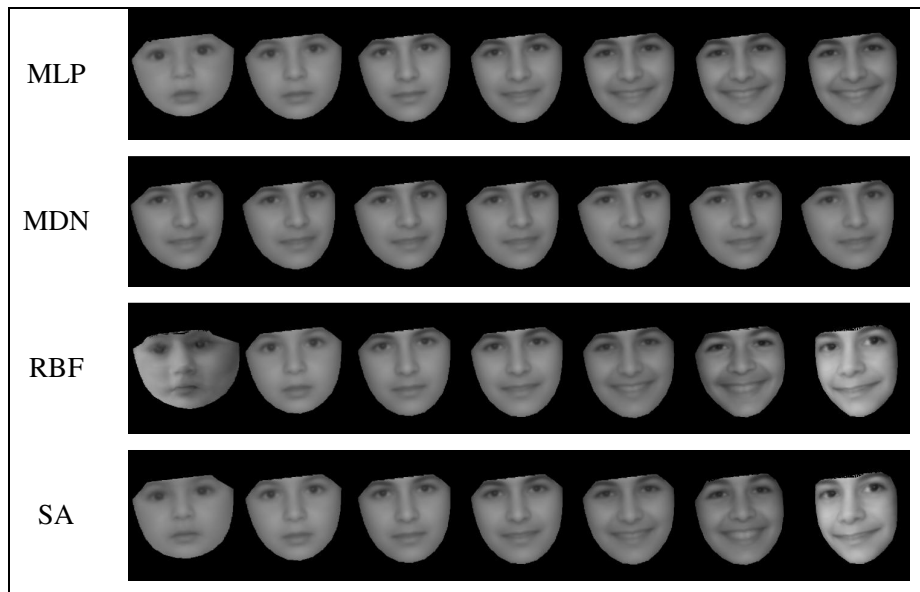


Figure 7: Variation of mouth width using MLP, MDN, RBF networks and SA method

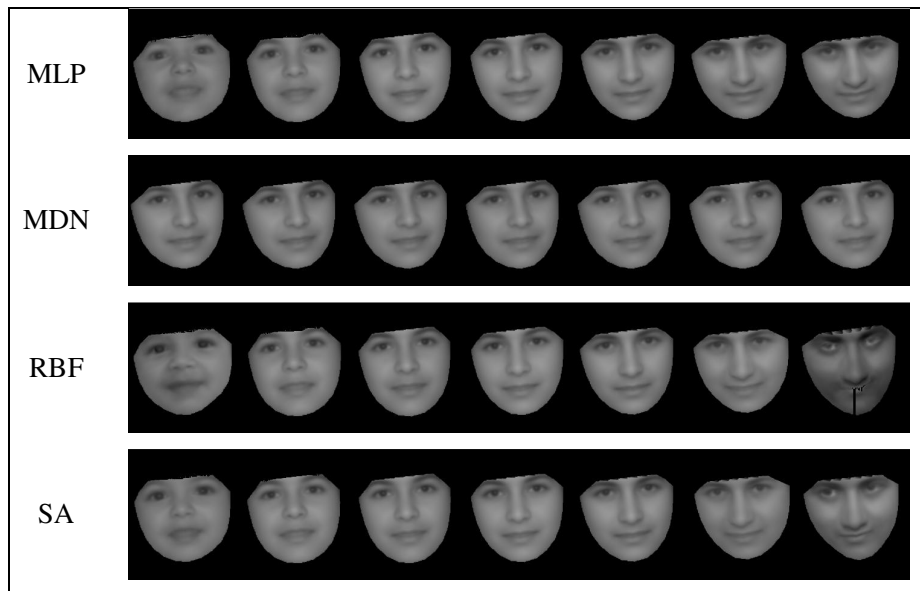


Figure 8: Variation of nose length using MLP, MDN, RBF networks and SA method

Based on the age of the subjects and the corresponding model parameters we establish a mapping, which gives us the set of model parameters corresponding to a face image at certain age. Since

the ages of subjects in the training set ranges from 0 to 69 years old, the corresponding mapping is suitable only for this age range. Figure 9 illustrates the age variation in the reconstructed face images based on the model parameters obtained with the MLP, MDN, RBF and SA methods for inputs corresponding to the ages 1, 5, 10, 15, 25, 30, 35, 40, 45, 50.

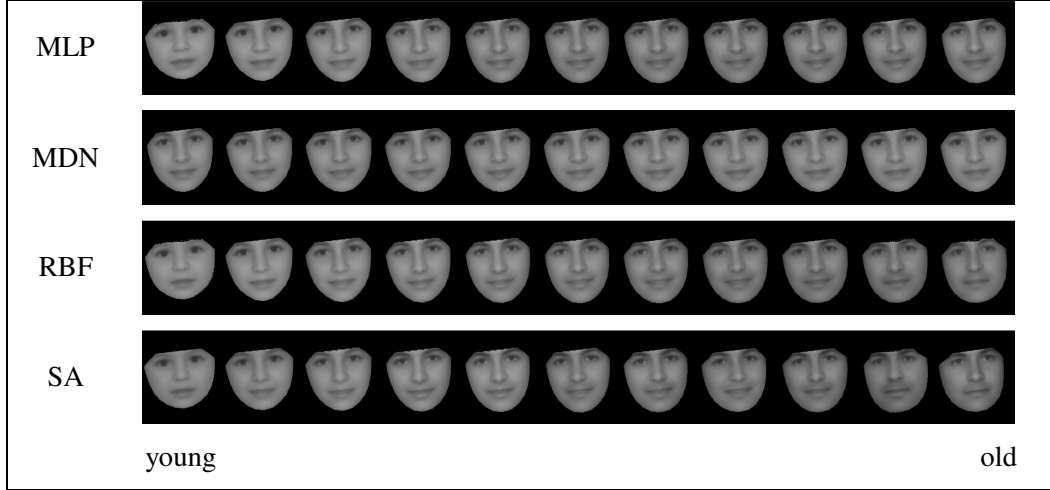


Figure 9: Variation of age using MLP, MDN, RBF networks and SA method

Figure 10 illustrates the variation of the latent variable in the reconstructed face images based on the model parameters obtained with the trained RBF mapping, which is part of the GTM model.



Figure 10: Latent variable rotation variation isolated using GTM

Both quantitative and visual results obtained with the MLP, RBF and SA methods clearly show the variation of the quantities for aspect ratio, distance between eyes, mouth width and nose length on the reconstructed face images. The results corresponding to the RBF model are better than those corresponding to the MLP and SA models. The results obtained with the MDN method do not show any variation of the quantities due to the large dimensionality of the problem. This could be explained with the fact that the modality of the distribution of the target data is very large and two mixture components are not sufficient to model this distribution. However, increasing the number of mixture components would lead to a problem with a very large number of outputs, which is not feasible to solve. Moreover we are interested in generating one specific solution of the multi-valued mapping, which gives a typical representation of the target data

distribution for a given input value of the source of variation. The visual results obtained with the GTM model presented in Figure 10 show variation in the rotational pose of the face images. It is true that this type of variation has the most substantial effect within the training set, thus the visual results obtained for the GTM models are reasonable. However, with the GTM method it is not possible to isolate a chosen attribute especially in the cases that the variation related to that attribute is not dramatic. In the problem that we consider, we are interested in defining the relationship between a specific variable of interest and the representation of the observed data, rather than finding the generating process of the observed data from a small unknown, in this case, number of latent variables. Therefore the use of the GTM method for this problem is not well suited.

4.3 Face Image Deformations Application

Once the mapping in equation (1) is established we can use it to generate typical face samples that display to a predefined extent the isolated quantity of interest. It is also possible to alter a certain facial attribute of an existing face using:

$$\mathbf{b}_{\text{new}} = \mathbf{b}_{\text{now}} + [\mathbf{g}(q_{\text{new}}) - \mathbf{g}(q_{\text{now}})] \quad (2)$$

where \mathbf{g} is the one-to-many mapping function, \mathbf{b}_{now} is the vector of model parameters for the source face image, q_{now} is the estimated value for an attribute in the source face image, q_{new} is the new value of the attribute, and \mathbf{b}_{new} is the estimated vector of model parameters for the new attribute which can be used as the basis for reconstructing the destination image (see Edwards et al., 1998 and Cootes, Edwards & Taylor, 2001 for details). In effect equation (2) modifies the current set of model parameters of an existing face image, in order to accommodate a change in a certain facial attribute. Given an image synthesized using the method described above, it is possible to calculate the intensity of a facial attribute and compare it with the pre-defined value of that attribute (q_{new}), enabling in this way the quantitative evaluation of the accuracy of the method in generating face images that display facial attributes of pre-specified strength.

To demonstrate the application of the proposed approach to face image deformations we use the trained model for the RBF network and the method described in the previous paragraph to generate face images that deform previously unseen faces from the test set (504 face images). For each of the attributes age, aspect ratio, distance between eyes, mouth width and nose length we select three values close to the minimum value, maximum value and average value of the respective attributes for the entire set of training images. The actual values that we used in the

experiments are given in Table 6. We then use equation 2 for obtaining model-based representations of deformed face images from the test set. Table 6 shows the quantitative results from these experiments showing the absolute mean errors and standard deviations of the quantities for aspect ratio, distance between eyes, mouth width and nose length on the reconstructed face images. The mean absolute error shows the mean difference between forced attribute values and the attribute values calculated from all reconstructed test faces (504 test cases).

Table 6: Quantitative results face image deformations experiments

Attribute Values	Mean absolute error	Standard deviation
Aspect Ratio		
0.6213	0.00465	0.00379
0.8911	0.00411	0.00330
1.1609	0.00739	0.00624
Mean error	0.00538	
Distance between eyes		
0.3711	0.00470	0.00575
0.4547	0.00571	0.01066
0.49	0.07164	0.01299
Mean error	0.02735	
Mouth width		
0.2253	0.00541	0.00411
0.404	0.00347	0.00293
0.5	0.19310	0.00718
Mean error	0.06733	
Nose length		
0.2390	0.00983	0.04165
0.3469	0.00742	0.03642
0.4	0.13983	0.00233
Mean error	0.05236	

Figure 11 shows five original face images from the test set and Figures 12, and 13 show visual results obtained from the face image deformation experiments for the age and mouth width attributes using the five face images shown in figure 7a.

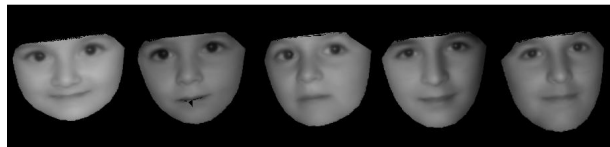


Figure 11: Original face images

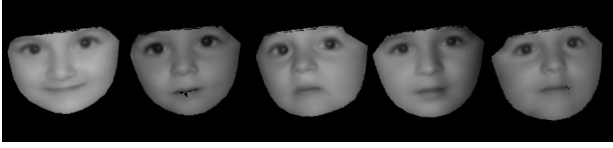


	Young age
	Middle age
	Old age

Figure 12: Deformations of face images due to age variation

	Min value
	Middle value
	Max value

Figure 13: Deformations of face images due to mouth width variation

Visual results display deformations associated with the modification in the facial attribute activated. The quantitative results in Table 6 demonstrate that errors between the values of the attributes chosen and the actual values on deformed images are small, proving in that way the viability of this approach.

A different approach to the face image deformations application based on a specific function that models the relationship between face model parameters and a specific facial attribute has been

used in (Lanitis, 2003). The advantage of the neural network approach described here is that it does not make an assumption about the mapping function and the neural network model is much simpler to update.

6. Conclusions

In this paper we investigate the use of a number of different techniques in the task of finding the mapping between a specific source of variation within a multivariate data distribution and the multivariate data distribution itself. The source of variation represents a quantity of interest related to the problem of face image synthesis. We look for such mapping which gives a typical representation of the face images distribution that exhibits the variation of the specific quantity of a facial feature. In this mapping, the target space has a higher dimension than the input space and for one input value the target output value is not unique. This leads to finding one-to-many multi-valued mapping. More specifically, we investigate several well-known methods used for solving such problems including MLP, RBF, MDN and GTM.

As part of the experimental evaluation the aforementioned techniques were applied to the problem of face image synthesis given a quantified description of a certain facial attribute such as aspect ratio, mouth width nose length and age. In this context the SA, MLP, RBFs and MDN are trained to map a specific facial attribute to model parameters corresponding to face images. The GTM algorithm is used for mapping one latent variable to the data space of face model parameters.

The results of the conducted experiments demonstrate the potential of using RBF's for isolating sources of variation and generating typical representations of the corresponding face image distributions. With the neural network approach no assumption is made about the mapping function; the neural networks are learning the complex mapping between the desired attributes and the parameters related the specific applications. The best performance is achieved with the RBF method, which could be attributed to the fact that in general RBFs are more appropriate for regression/interpolation type of problems. The MLP and RBF methods give the conditional averages of the target data conditioned on the input vectors and as expected they do not give a complete description of the target data (Bishop, 1994) , (Richmond, 2001). In this problem it is sufficient to define a mapping that generates typical samples of the data distribution (and not its entire variance) given specific values of the desired source of variation. This makes the results

with the MLP and RBF (which are relatively simple methods, compared to MDN and GTM) sufficiently good for the type of inversion problems that are addressed, compared to the MLP results for the acoustic-to-articulatory inversion mapping reported in (Carreira-Perpiñán, 2001), (Richmond, 2001). For this one-to-many problem and also the problem of reconstructing the same spectrum from different spectral line parameters (Bishop, 1994), the entire variance of the distribution is required. It has to be noted also that for the one-to-many problems considered in this paper the training algorithm for the MLP does not have to be modified as suggested by Brouwer (2004), resulting in increased complexity.

While the MLP and RBF methods provide continuous mappings, the SA method gives only a discrete mapping of an interval to a point in the target space. As the number of intervals in the SA method increases we are moving towards the actual data. Therefore this method is not expected to have the generalisation capabilities provided by the neural network models.

The MDN (Bishop, 1995) can give a complete description of the target data conditioned on the input vector provided that the number of mixture components is at least equal to the maximum number of branches of the mapping. The experiments carried out in the considered application demonstrate that in problems for which the modality of the distribution of the target data is very large and not known, the application of the MDN leads to a large number of mixture components and outputs. Therefore it does not provide the desired type of mapping, which explains the poor results obtained with the MDN method. In particular, these results do not show the desired variation of the quantities of interest.

The GTM (Bishop, Svensén & Williams, 1998) is used to map points in the latent variable interval to points in the target data space and our results with this method actually show this mapping of one latent variable to the corresponding data spaces. It has to be noted though, that in order to isolate a specific source of variation, we need to find all latent variables which leads to a problem that is not computationally feasible (Cootes, Edwards & Taylor, 2001). Due to the large number of possible sources of variation and the exponential computational growth, we can not isolate a specific quantity of interest as required.

The framework presented in the paper can be potentially useful in various applications involving multivariate distributions. For example in the face application, the definition of a mapping between a type of facial variation and facial appearance can prove useful in applications

involving the synthesis of face images displaying a particular facial attribute. An initial attempt towards this direction is presented in section 4.3. In this context, the RBF neural network model was used as the basis for altering certain facial attributes of existing faces and obtain face image deformations that exhibit these facial attributes. Similar techniques can be used in forensic sketching where faces displaying certain attributes are constructed in an attempt to synthesize suspect's mug shots. Also facial animation and caricaturing algorithms can rely on mapping functions that relate different sources of facial variability with facial appearance.

Future work in this area will focus on two distinct directions: (i) investigation of the performance of other methods and (ii) applications to other problem domains. Although the performance achieved by two of the methods RBF's is satisfactory, the use of other types of methods that could be applied to the problem will be considered. For example we are in the process of evaluating the use of Support Vector Machine Regressor (Vapnik, 1979, 1995) for modelling the relationship between a parameter and a multivariate distribution.

References

- Bartholomew, D.J. (1987). *Latent Variable Models and Factor Analysis*, London: Charles Griffin & Company Ltd.
- Belhumeur, P. N., Hespanha, J., Kiregeman, D. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7), 711-720.
- Bishop, C.M. (1994). Mixture Density Networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University. Retrieved online from <http://research.microsoft.com/~cmbishop/downloads/Bishop-NCRG-94-004.ps>, Last Accessed December 2011.
- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- Bishop, C.M., Svensén, M., & Williams, C. K. I. (1998). GTM: The Generative Topographic Mapping. *Neural Computation*, 10(1), 215-234.
- Brouwer, R. K. (2004). Feed-forward neural network for one-to-many mappings using fuzzy sets. *Neurocomputing*, 57, 345-360.
- Carreira- Perpiñán, M. A., (1999). One-to-many mappings, continuity constraints and latent variable models. Proceedings IEE Colloquium on Applied Statistical Pattern Recognition. (pp. 14/1-14/6). Birmingham, UK.
- Carreira-Perpiñán, M. Á. (2001). Continuous latent variable models for dimensionality reduction

and sequential data reconstruction. PhD thesis, Dept. of Computer Science, University of Sheffield, UK (Retrieved online from <http://faculty.ucmerced.edu/mcarreira-perpinan/papers/phd-thesis.html>, Last Accessed November. 2011).

Cootes, T. F., Edwards, G.J., & Taylor, C.J. (2001). Active Appearance Models. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 23, 681-685.

Dempster, A., Laird, N., & Rubin. D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39(1), 1-38.

Duda, R. O., Hart P. E. (1973). *Pattern Classification and Scene Analysis*. New York:Wiley.

Edwards, G.J., Lanitis, A., Taylor, C.J., & Cootes, T.F. (1998). Statistical Face Models: Improving Specificity. *Image and Vision Computing*, 16(3), 203-211.

Jayne, C., Lanitis, A., Christodoulou, C. (2011). Neural network methods for one-to-many multi-valued mapping problems, *Neural Computing and Applications*, 20(6), 775-785.

Lanitis, A. (2003). PROSOPO-A Face Image Synthesis System. In Y. Manolopoulos, S. Evripidou, and A. Kakas (Eds.), *Advances in Informatics. Post-proceedings of the 8th Panhellenic Conference in Informatics. LNCS 2563*. (pp. 297–315). Berlin: Springer.

Lanitis, A. (2008). Comparative Evaluation of Automatic Age Progression Methodologies, *EURASIP Journal on Advances in Signal Processing*, Article ID 239480 (10 pages).

Lanitis, A., Taylor, C. J., & Cootes, T. F. (1997). Automatic Interpretation and Coding of Face Images using Flexible Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 743-756.

MacKay, D. J. C., & Gibbs, M. N. (1998). Density networks. In *Statistics and Neural Networks: Advances at the Interface*, eds. Kay J. W. and Titterton D. M., Oxford: Oxford University Press, pp 129-146.

Møller, M., (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4), 525-533.

Powell, M.J.D. (1985). Radial basis functions for multivariable interpolation: A review. *IMA Conference on Algorithms for the approximation of Functions and Data* (pp. 143-167). RMCS, Shrivenham, England.

Richmond, K. (2001). Mixture Density Networks, Human articulatory data and acoustic-to-articulatory inversion of continuous speech. Retrieved online from http://www.cstr.ed.ac.uk/downloads/publications/2001/Richmond_2001_a.ps (Last Accessed November. 2011)

Rumelhart, D.E., Hinton D.E., & Williams, R. J. (1986). Learning representations by back-propagation errors. *Nature*, 323, 533-536.

- Vapnik, V. (1979). *Estimation of Dependences Based on Empirical Data [in Russian]*, Nauka: Moscow. (1982). English Translation: New York: Springer Verlag.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag: New York.
- Zeng, X. & Yeung D. S. (2003). A Quantified Sensitivity Measure for Multilayer Perceptron to Input Perturbation. *Neural Computation*, 15, 183-212.