Computer Vision and Image Understanding 116 (2012) 1022–1031

Two dimensional synthetic face generation and verification using set estimation technique ☆

Madhura Datta [a], C.A. Murthy [b,*]

[a] UGC-Academic Staff College, University of Calcutta, 92 Acharya Prafulla Chandra Road, Kolkata 700 009, India
[b] Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India

ARTICLE INFO

Article history:
Received 29 June 2010
Accepted 16 May 2012
Available online 6 June 2012

Keywords:
Feature extraction
Minimal spanning tree
Intra-class threshold
Set estimation
Inter class feature and intra class feature
PSNR

ABSTRACT

In this paper set estimation technique is applied for generation of 2D face images. The synthesis is done on the basis of inheriting features from inter and intra face classes in face space. Face images without artifacts and expressions are transformed to images with artifacts and expressions with the help of the developed methods. Most of the test images are generated using the proposed method. The measured PSNR values for the generated faces with respect to the training faces reflect the well accepted quality of the generated images. The generated faces are also classified properly to their respective face classes using nearest neighbor classifier. Validation of the method is demonstrated on AR and FIA datasets. Classification accuracy is increased when the new generated faces are added to the training set.

1. Introduction

Synthetic face generation for the purpose of face recognition has been explored in recent times. 2-Dimensional (2D)–3-dimensional (3D) reconstruction and generation of new faces with various shapes and appearances received attention of the biometric researchers. The popular solution to the problem is proposed by 2D and 3D modeling of faces. 3D models include face mesh frames, morphable models and depth map based models, where one needs to incorporate high quality graphics and complex animation algorithms. Flynn et al. [1] provided a survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, Tao et al. [2] derived 3D head poses from 2D to 3D feature correspondences. Blanz et al. [3] proposed face recognition based on fitting a 3D morphable model with statistical texture. Four main approaches for 2D modeling are active appearance models (AAMs) [4], manifolds [5], geometry driven face synthesis methods [6] including face animation [7], and expression mapping techniques [8–10].

2D to 2D face reconstruction was first developed by Cootes et al. [4] and their active appearance model generated from the 2D face images is one of the powerful methods. Multi-view face reconstruction in 2D space is done by manifold analysis [5]. Geometry driven face synthesis by Zhang et al. [8], and expression mapping techniques are also useful in 2D face generation [11]. Table 1 provides comparative appraisal on advantages and disadvantages of some useful techniques of face synthesis. All the methods mentioned in Table 1, new face images are generated either from a model or from some functional properties.

In this paper the idea of developing new face images is based on statistical properties. A global formation of a face class for a particular person P can be assumed to be a set consisting of infinitely many face images. Face images of P differ from each other mainly because of muscle movements in different portions of the face. Different muscle movements result in different expressions. Movement of eyebrows, twitching of nose, muscle movement in cheeks, movement of lips, opening and closing of mouth along with different combinations these are some examples of movements in different parts of face.

If his/her face is continuously photographed, one may also note that, for any person, there exists a path joining those two images. Theoretically, a set $A \subseteq \mathcal{R}^m$ is said to be path connected, if for any two points $x, y \in A$, $\exists$ a continuous function f from [0, 1] to A such that $f(0) = x$, $f(1) = y$. That is, there exists a path containing infinitely many images joining x and y. Path connectivity is a valid assumption, since disconnected set of face images of the same person cannot occur. Therefore, we may mathematically define the face class as a path connected set. If we represent an image of a particular expression in the face of a person P by a vector x_0, then the set corresponding to the small variations in the same expression may be assumed to be a disc of radius $\varepsilon > 0$ around x_0. Using the set estimation principles, the face class for the person is estimated and the value of the radius $\varepsilon > 0$ from the estimated

☆ This paper has been recommended for acceptance by Martin Levine.
* Corresponding author.
E-mail address: murthy@isical.ac.in (C.A. Murthy).

Table 1
Comparative advantages/disadvantages of 2D to 2D face synthesis with respect to some selected papers.

	Method	Advantages	Disadvantages
Cootes et al. [4]	Uses AAM and ASM models. Statistical texture and shape analysis techniques are used	Results are satisfactory	The initial feature points on face images are manually annotated in training stage
Huang et al. [5]	Faces in Manifold are reconstructed with expressions	Expressions are generated in neutral faces	Many face images of a person taken from different angles are needed for the construction of a manifold
Liu et al. [6] (ERI algorithm)	Expression ratio images are reconstructed	Applications are done on 2D faces and expressions are added successfully.	The method can only map expressions present in the probe image. Smooth transition from one face image to the other is not possible
Zhang et al.[8]	Facial expressions were synthesized feature point set	Expression editing software is available	Additional method is required for the generation of feature point set
Pyun et al. [12]	In this expression mapping technique, geometry controlled image warping method is used	Morphing can be done easily because of the knowledge of geometry of face	Geometrical properties of the face under consideration are to be found and stored
Pighin et al. [13]	Basis expression space is created for every person in the training set	Expressions of another person can be inherited	Construction of a person's expression space needs prior computations and also large memory space
Neely et al. [14]	Morphing operators are used for the construction of new faces	Easy to implement	Limited number of expressions are generated. Smooth transition from one face image to the other is not possible

set can be obtained. Inside this radius ε, say in m-dimensional Euclidean space, many different face images can be generated for the same person.

The method of set estimation is mainly used to find the pattern class and its multi-valued shape/boundary from its sample points in $\mathcal{R}^2$ [15–20]. Some investigation on estimation of α-hull for point sets in $\Re^3$ had been proposed by Edelsbrunner et al. [18]. Later Mandal et al. [19] extended the method to higher dimensions and found it useful in developing multi-valued recognition systems. As one can obtain the shape or boundary of a given finite set of points, the procedure of set estimation also generates the intuition for determination the radius of the disc. As the number of face images of the person increases, we shall be obtaining more information regarding the face class and hence the radius value needs to be decreased. Thus, the radius value is a function of the number of images and it is independent of the center of the disc. As a tool of set estimation, minimal spanning tree (MST) is used to calculate disc radius for each class. One of the methods of calculating the radius is (i) find MST of the set of n points with the edge weight as Euclidean distance, (ii) take L_n as the sum of the edge weights of MST and (iii) take the radius as, $\sqrt{l_n/n}$.

In contrast, in this paper, alternative procedure of using MST is adopted. Two algorithms are designed here to generate the new face images from the estimated sets. In the first method, new face images of a particular person are generated using features of that class only. In the second one, new face images are generated using features not only of that class, but also the features from other classes.

The face data bases used for experiment are AR [21] and FIA [22] databases. Both the AR and FIA data sets consist of frontal faces. In AR database, faces have illumination and artifact variations, while in the FIA database, the face images have varying facial expressions. Both the datasets have suitable properties for designing new morphed faces which are helpful for recognition purpose.

The paper consists of eight sections. Section 2 consists of brief statistical overview of set estimation technique and Section 3 introduces minimal spanning tree as set estimation tool. Two algorithms are designed for face generation in Section 4. Sections 5 and 6 consist of analysis of experimental results. In Section 7 classification accuracies of the method are discussed. Section 8 is the concluding section.

2. Brief statistical overview of set estimation for face classes

Before discussing the basic intuition for the applied method we assume that the dimensionality reduction technique has been already done using principal component analysis (PCA) [23,24]. Using PCA the image space is converted to face space. The whole process of set estimation, face class formation, and all the designed algorithms are carried out in reduced dimension, i.e. $\mathcal{R}^m$. Let a closed disc of radius r and centered at x_0 be represented by $\upsilon(x_0, r)$. That is, $\upsilon(x_0, r) = \{x : d(x_0, x) \leqslant r\}$, where d denotes the Euclidean distance.

If we represent a face image of an expression of a person P by a vector x_0, then the set corresponding to the small variations in the same expression may be assumed to be a disc of radius $\varepsilon > 0$ around x_0. The set corresponding to an expression of the person P may be taken as $\bigcup_{i=1}^{N}\{x \in \Re^m : d(x_i, x) \leqslant \varepsilon\}$, where $x_1, x_2, \ldots, x_N$ are N vectors corresponding to N images of the same expression for the same person. The dimension of the vectors is assumed to be m. The set corresponding to the union of all possible expressions of a person may also be taken as a connected set. The face class of a person is nothing but the set of all possible face images of that person.

In the above formulation, as the number of given face images of the person increases, we obtain more information regarding the face class and hence the radius value decreases. Thus, the radius value is a function of the number of images. Usually one may want to "estimate" a set on the basis of the given finitely many points. Grenander has formulated the set estimation problem as the problem of finding consistent estimate of a set [17].

Let α_n be an estimated set based upon the random vectors $X_1, X_2, \ldots, X_n$. Then α_n is said to be a consistent estimate of α, if $E_\alpha[\mu(\alpha_n \Delta \alpha)] \to 0$ as $n \to \infty$, where Δ denotes symmetric difference, μ is the Lebesgue measure and E_α denotes the "expectation" taken under α. Then

$$\alpha_n = \bigcup_{i=1}^{n}\{x \in \Re^2 : d(x, X_i) \leqslant \varepsilon_n\} \tag{1}$$

The discussed theorem did not mention a way of finding ε_n. Murthy [15] developed a way of finding ε_n for points in 2D spaces. He also generalized the method to any continuous density function on α, where α is a path connected set. His method has applications in different fields and a few of its modifications are all documented in literature [15–20]. None of these methods deals with generating new face images.

3. Minimal spanning tree as a tool of set estimation method [16]

Initially face space is created by any dimensionality reduction technique which is followed by the set estimation method. The next task is to make the estimated set path connect, where the connectivity is preserved by MST. The steps of a generic way of making the estimated set path connected is described below, where only finite union of disks is considered.

(a) Find minimal spanning tree (MST) of $S = \{X_1, X_2, \ldots, X_n\}$, where the edge weight is taken to be the Euclidean distance between two points and S is a finite set of points.
(b) Take ε_n as maximum of the (n-1) edge weights of MST.
(c) Take the estimate A_n as

$$A_n = \bigcup_{x \in S} \{y : d(x, y) \leqslant \varepsilon_n\} \quad (2)$$

where ε_n is not dependent on the dimension m and A_n is now path connected.

To apply the above method in face analysis, several face images for the same human being are considered. The number of classes is same as the number of human beings. Let us consider M human beings and for each human being we have N face images of same size and same background. If we represent an image by a vector $\underline{x}$, then we are considering all possible such vectors corresponding to a human being as a set represented by α. This set denotes the face class of that human being. It can be noted that we do not know α completely. Only a few points of α like the different expressions of a face are known to us and considered as training set. Initially we apply PCA to reduce the number of dimensions to m. Thus every face now is an m dimensional vector.

3.1. Determining the value of radius (ξ_i) of a face class

For each class we calculate MST of the respective N vectors and find its maximal edge weight. Let us denote the maximal edge weight of the MST of the ith class by ξ_i. It is found in the following way. If the reduced set for the ith class after dimensionality reduction is denoted by $\{z_{1i}, z_{2i}, \ldots, z_{Ni}\}$, then the estimated set for the class i, is denoted by B_i, and is obtained as

$$B_i = \bigcup_{j=1}^{N} \{y \in \Re^m : d(y, z_{ji}) \leqslant \xi_i\}, \quad i = 1, 2, \ldots, M \quad (3)$$

where ξ_i is defined to be the maximal edge weight of the minimal spanning tree (MST) for each class i.

4. Proposed algorithms for generating synthetic face points in the face space

Two algorithms have been designed to implement the idea of generating new face points in face space. Then, the face images corresponding to the newly generated face points are reconstructed. These new face images are expected to provide information combined from the two face points. If the two face points belong to the same class, then they are expected to provide mixed information corresponding to the original two points. On the other hand, if the two points are from two different classes, then they are expected to provide mixed of information of the two classes. Sometimes, the mixed information, either from the two points belonging to same class or from the two points belonging to different classes, may distort the original face images. If the two discs corresponding to the two points do not intersect, then the mixed information may become blurred. On the other hand, if the two discs intersect, then the points belonging to the intersecting region are seen to possess, on many occasions, the mixed information. The algorithms 1 and 2 reflect the ideas as stated.

4.1. Generating face points taking the features from intra face classes in the space

4.1.1. Algorithm 1

This algorithm generates face points inside the face class following the estimated set and path connected properties of classes.

Step1: Find MST for each face class i.
Step 2: For MST of a face class and for each edge joining two points x and y, find intermediary points p, using the following equation

$$p = \lambda x + (1 - \lambda) y, \quad \lambda \in (0, 1). \quad (4)$$

Step3: Reconstruct the face corresponding to every new face point p generated.

The number of intermediary points and the corresponding values for λ are decided on the basis of the requirements. The process of selection of new face points is illustrated in Fig. 1.

In the above model we have two classes each having 5 probe face points denoted by and following the path of the MST, the intermediate face points denoted by are generated having the features of both the nodes of edge.

4.2. Generating face points taking the features from inter face classes in the space

The purpose of this section is to generate new face images for a face class which possesses features of face images from other classes. For example, the person in one face class may never wear spectacles, whereas in the other face class, the corresponding person wears spectacles. Another example is that the eyes of a person in one class may be open, where as the eyes of a person in another face class may be closed. The problem may be tackled in two ways. One way is to identify features which we want to see in a face image of a person and find another person who has the said features. Then make a combination of the features of both persons at the same location and judge whether the combination is meaningful. The second way is to take combinations of face images and decide whether the changes of features incorporated in the generated face image is meaningful (i.e., not a noisy image). In both the approaches, combinations of faces need be considered and judgments about the combinations are to be made.

In the first approach, the aim is simpler but is guided by heuristic choice and more computations are required. In the second case, new features are given more importance as decided by the algorithm. Therefore, we shall follow the second approach.

4.2.1. Algorithm 2

This algorithm generates face points taking the features from other face classes in the space.

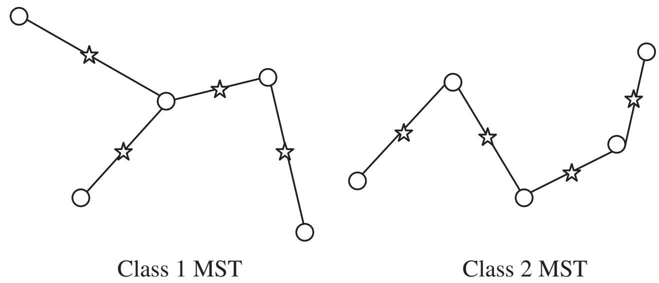

Fig. 1. Intra class facial feature generation.

Step1: Find ξ_i for class i. using the method described in Section 3.

Step 2: For every class i, find its nearest face class j using the following steps.

Let $a_{ik} = Min\ d(z_{li}, z_{wk}),\ 1 \leqslant l,\ w \leqslant N,\ 1 \leqslant i,\ k \leqslant M$ and let $a_{ij} = Min_{k \neq i} a_{ik}$ Let, without loss of generality, $z_1 \in \{z_{1i}, z_{2i}, \ldots, z_{Ni}\}$, $z_2 \in \{z_{ij}, z_{2j}, \ldots, z_{Nj}\}$ be such that $a_{ij} = d(z_1, z_2)$.

Step3: If $a_{ij} \geqslant \xi_i + \xi_j$, then we shall go to the next value of i. Otherwise, join the two points z_1 and z_2 by a line segment and generate a face point from the intersecting region of the discs of radii ξ_i and ξ_j centered at z_1 and z_2 respectively and falling on this line segment.

Step3.1: The geometrical formulation used to generate points on the line segment joining the points A and B defined by

$$A = \frac{(a_{ij} - \xi_j)z_2 + \xi_j z_1}{a_{ij}}; \quad B = \frac{(a_{ij} - \xi_i)z_1 + \xi_i z_2}{a_{ij}} \tag{5}$$

where $d(z_1, z_2) = a_{ij}$, $d(z_2, A) = \xi_j$, $d(z_1, B) = \xi_i$, $d(z_1, A) = a_{ij} - \xi_j$ and d is the Euclidian distance.

Step 4: Reconstruct the new faces corresponding to the generated face points.

The above mentioned linear combinations are used to generate the face points on the line segment.

In Fig. 2, two face classes are formed in the face space with two intersecting disks with radii ξ_1 and ξ_2. New face images are generated from the line portion '▬▬' of the intersection region A and B. The line joining the probe disks are centered at z_1 and z_2. The generated face points must incorporate features from two different classes. In reality, these features can also be the artifacts used in faces.

5. Experimental design and result analysis

5.1. Dataset description

For performing experiments, both the AR and FIA dataset are organized in such a way so as to justify the proposed method. For inter and intra class feature sharing algorithms every dataset are divided into training and test parts. These datasets are chosen for the variations of images having artifacts like the mufflers, sunglasses, etc. in AR set and faces with expressions like happiness, twitching of eyes, movement of mouth portions in FIA dataset. It is assumed that the faces without any variations (neutral faces) in the training sets will inherit features from the test faces having artifacts and other expressions.

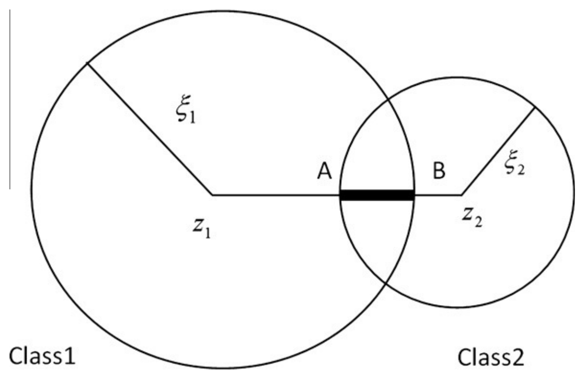

Fig. 2. Inter class facial feature generation.

For AR dataset 40 classes are considered from which five face images of each class are selected as training images. Five training set images correspond to (i) one neutral face, (ii) one smiling face, (iii) one face with muffler, and (iv) two faces with sunglasses. Size of the images is cropped to 35×46 and the background of each image is white. Other parts like collars of shirts, hair are present in the images. For generating new face images the considered value for λ = 0.5 and 4 new images for each class have been generated. For FIA dataset all 13 classes are considered and five faces are chosen as training set. The facial properties for the training set are one neutral face and four faces with expressions. Number of generated faces is seven for each edge in MST. The value of λ for the experiment are λ = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8. The input images are of size 64×64 and do not have any background.

For construction of MST we have used 5 face points for training from both the datasets. PCA has been used for reducing dimension and the whole process of generating faces is carried out with reduced dimension. The dimensions after reduction for AR and FIA datasets are 200 and 65 respectively. The experiments are carried out in two different ways and three resultant sets are obtained. Result sets 1 and 2 show the generated faces after applying algorithm 1 on AR and FIA datasets respectively. Result sets 3 and 4 show the results of applying algorithm 2 on AR dataset.

5.2. Face generation using algorithm 1 (with intra class features)

The probe set (training set) is constructed with five face images from each class. Subspace is formed with the help of PCA based dimensionality reduction method. In the face space, for each class, MST is constructed with the face points of the probe set. New faces are generated using the method stated in the algorithm 1.

Algorithm 1 is applied on the training set. MST for each class would have four edges. Thus, four new images are generated for the same class. The total number of images that could be generated is 160. Because of constrains of space, some representative generated face images are shown in the column 3 of Fig. 3. These generated face images are sharing the properties of the images in columns 1 and 2.

5.2.1. Results on PSNR test and remarks on dataset 1

PSNR (Peak signal to noise ratio) is used to judge the quality of an image with respect to a given image. A PSNR value is greater than 30 dB, in general, indicates the closeness of the images. Quality of the generated image is taken to be acceptable if the PSNR is found to be greater than 20 dB [24]. 160 new images, taken four images per class from 40 classes have been generated. PSNR value of each one is found to be greater than 35 dB. In Fig. 3, seven among the 160 generated images are shown.

The remarks on dataset 1 are given as:

(1) In the generated images of rows 1–5, the eyes are visible from spectacles. Thus, the effect of sunglasses is not present in the generated images.
(2) Illumination is adjusted in the generated faces. In rows 1, 3, and 5 the first image is left illuminated, whereas the second one is right illuminated. The resultant images have no pronounced left or right illuminations.
(3) In rows 6 and 7, the generated face images inherited sun glasses (one feature) from one face and muffler (other feature) from the other face. Generated faces possess both the artifacts used by the two persons from which the face is generated.
(4) Note that the number of new images created for two different classes are same. This is because of (i) the number of training sample points from each class is 5, (ii) the MST has four edges for each class and (iii) one image is generated per each edge of MST.

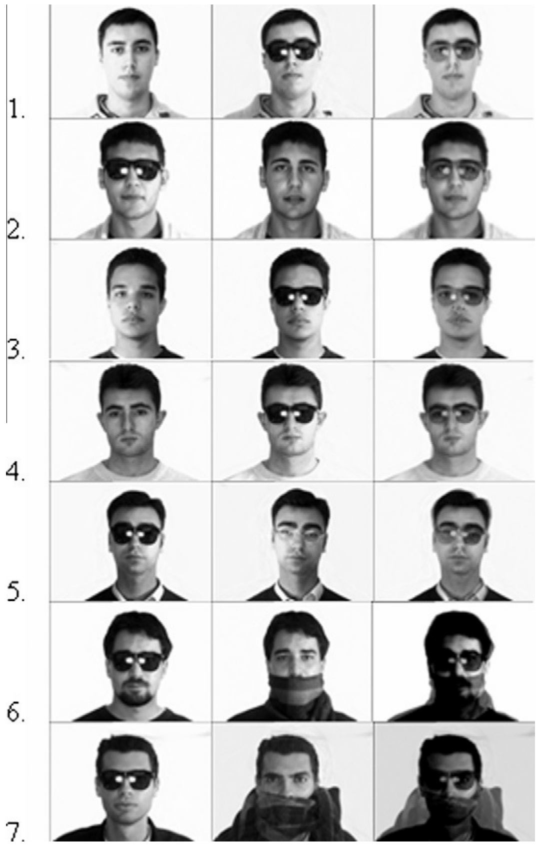

Fig. 3. Resultant set 1. For each row, the image in the third column is generated as a synthetic image from the images of columns 1 and 2.

5.2.2. Test on dataset 2 and remarks

Algorithm 1 is applied on the training set of FIA dataset. The purpose of this experiment is to show the smooth transition from one expression to another. For each edge in MST, seven images are generated by taking the values of λ as 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8. Seven such results are shown in Fig. 4.

In each row of images in Fig. 4, the first and the last images are the input images and the rest are generated images. The generated faces correspond to the intermediate face points of the edges joining the two end-training points of the MST. These sets of images can also produce a video showing the change of facial expressions.

5.3. Face generation with inter class features using algorithm 2

In the experimental part the generated images having the dominant properties of the training set but some addition face properties like artifacts or expressions are inherited in the generated faces. Two sets of images are shown in Figs. 5 and 6. Both sets of images are generated using algorithm 2, where the features of two intersecting classes are mixed. And the newly generated faces are the convex combinations of the intersecting classes. All the generated faces are from the intersecting region and on the line joining the points A and B shown in Fig. 2. In image set 2 the smooth changes of two different faces of one person to another can be clearly visualized. The experimental dataset is generated from AR dataset with training set of neutral faces without any artifacts. The corresponding indices for the faces are 1, 2, 3, 4 and 5. Number of faces in test set consist of all images from other classes.

This way of dataset formation is useful to show the efficiency of the algorithm. The experimental result sets shown in Figs. 5 and 6 are applications on AR dataset. Similar results on FIA data set for expressions can be obtained.

Algorithm 2 is applied on face images of two different classes (Fig. 5). In each row, the first image is of the person whose characteristics are to be transferred. The second image is of the person receiving these characteristics. The third image is the image of the second person after receiving the said characteristics. Thus, the third image is the generated face from the first two images. The generated image corresponds to the middle point of the line segment falling in the intersection region, and it joins the centers of the two discs.

5.3.1. Result of PSNR test and remarks

The PSNR values for the generated images are computed with respect to the original images to show the quantitative outcomes on the quality of the generated images. All the PSNR values range from 25 dB to 35 dB. From the PSNR values, the second image is found to be closer to the generated image than the first. The number of such intersections between the classes is not more than 15. Here we provided 11 such instances. Thus, the PSNR values justify the generation of a face from two different classes, though intersecting.

The remarks are given as,

(1) As found from the images, one can view that the spectacles are added in each of the morphed faces of rows 1–8. The sunglasses are the dominant features added to the face images.
(2) In rows 9–11, the column 1 face images are “mufflered” faces, each generated face (column 3) image has a muffler. Note that the hidden portion of the face can be viewed in the generated face.
(3) It is not possible to generate new images for every pair of face classes. Non intersecting classes need not generate meaningful new faces. Thus the number of new images added to two different face classes in the data base may not be same.
(4) It is apparent from the images of column 3 that, in every row, the generated face is the face of the person shown in column 2 but not the one shown in column 1. To find mathematically the class identity of the generated faces, the Euclidean distance has been calculated and 1 NN classifier is used to classify each generated face. Since our intention is to add artifacts to the images of the column 2, we expect the generated set to belong to the class of the image in the column 2. In 100% cases the images are found to be classified to the proper class (i.e., the class of image 2 in each row).

One may also be interested in seeing the intermediate images when the face of one person is changed to the face of another person. The resultant set 4, shown in Fig. 6, depicts such transition for five cases. Similar results may be obtained in other cases too.

5.4. Rejection of the non meaningful face and corresponding PSNR test

Any linear combination of the two face images will not provide meaningful face images. Algorithm 2 has been developed with the notion that the intersecting classes only can share the facial features. Some rejected faces which were generated from the non intersecting classes are shown in the Fig. 7. From the subjective judgment it can be stated that these images are highly distorted and defining the face class is quite impossible.

There are many examples of such non-intersecting face classes. A few of them are shown in Fig. 7. In everyone of such cases, the PSNR value is found to be less than 10 dB. This again justifies the

Fig. 4. Resultant set 2. For each row, the smooth changes of images from one end point of the edge of the MST to another are shown.

contention that new face images should not be generated from non intersecting face classes.

6. Generalization capability of set estimation method: generating face images not in training set

In the previous sections, we have generated new images in between two training images, where those two images may be either from the same class or from different classes. In this section, we generate test images from the training images. Consequently, several intermediate images between a training image and a test image are generated. This will reflect the generalization capability of the method of generation of images using set estimation technique.

The flavor of the method of set estimation is that one can generate uncountably many images under a specific radius. If x is a vector in n dimensional Euclidean space, and y is a vector whose Euclidean distance from x is r, then y can be exactly generated from x by suitably using $(n-1)$ angle variables $\theta_1, \theta_2, \ldots, \theta_{n-1}$, where each $\theta_i \in [0, 2\pi]$. In fact, if $\theta_1, \theta_2, \ldots, \theta_{n-1}$ are chosen appropriately, then

$$y = x + r\begin{pmatrix} \sin\theta_{n-1}\sin\theta_{n-2}\ldots\sin\theta_1 \\ \cos\theta_{n-1}\sin\theta_{n-2}\sin\theta_{n-3}\ldots\sin\theta_1 \\ \cos\theta_{n-2}\sin\theta_{n-3}\sin\theta_{n-4}\ldots\sin\theta_1 \\ \cos\theta_{n-3}\sin\theta_{n-4}\sin\theta_{n-5}\ldots\sin\theta_1 \\ \vdots \\ \cos\theta_2\sin\theta_1 \\ \cos\theta_1 \end{pmatrix} \tag{6}$$

Additionally, many images can be generated corresponding to points on the line segment joining x and y. These images reflect the smooth transition from x to y.

After the reduction in the number of features, if the nearest neighbor of a test image y is x, then y is classified to the class of x (with the help of the nearest neighbor classifier). Let ξ denote the value of the threshold, calculated using set estimation

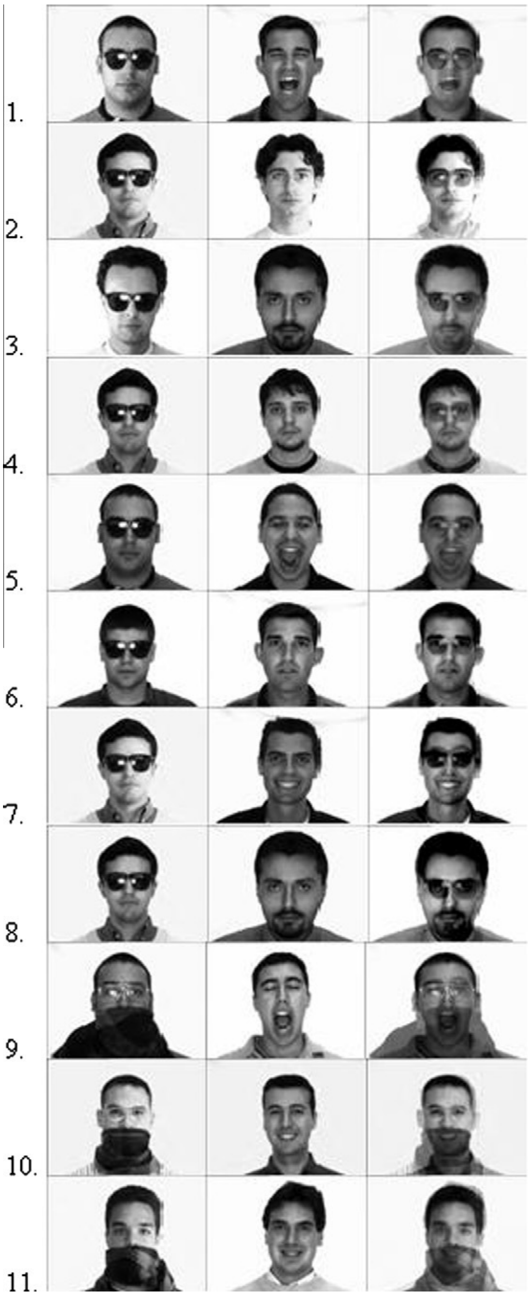

Fig. 5. Resultant set 3. For each row, the image in the third column is generated as a synthetic image from the images of columns 1 and 2.

procedure for the class of x and the distance between x and y be γ. Then, one of the following cases may arise.

(i) The classes of x and y are same and $\gamma \leqslant \xi$.
(ii) The classes of x and y are same and $\gamma > \xi$.
(iii) The classes of x and y are different and $\gamma \leqslant \xi$.
(iv) The classes of x and y are different and $\gamma > \xi$.

Case (i) occurs to at least 90% of test images. Since $\gamma \leqslant \xi$, y falls in the disc of radius ξ, and thus y can be exactly generated from x. We have shown 8 more equispaced synthetic images generated by varying the distance from x from 0 to γ without changing the values of θ_i s. These images are present neither in the training set nor in the test set of the data. PSNR values with respect to y increased to 35 dB as the distance from x is increased to γ. Fig. 8 shows the training image x, test image y and the intermediate images. The images are generated using Eq. (6).

Case (ii) occurs when $\gamma > \xi$, y does not fall in the closed disc of radius ξ centered at x. Thus, one can generate an approximation of y lying on the boundary of the disc and the distance between the approximated image and the actual image is the least among all images lying in the disc. Since $\gamma > \xi$, the generated images are not expected to be as close to y compared to the images generated in the previous case. In Fig. 9, some examples of x, y and the closest approximation to y are provided. The PSNR values with respect to y are found to be around 15 dB for the synthetic images.

Case (iii) does not usually occur, since y and x are from different classes as well as $\gamma \leqslant \xi$. Additionally, the occurrence of this case signifies that there are some common characteristics between y and x. Here y can be exactly generated from x. Among the considered data sets, this case did not occur even once.

Case (iv) conditions occur when a point is forcibly classified to a class and the information from the training set is insufficient for classification. In this case y is misclassified and y is less similar to points from its own class than from another class. This represents a case of training set being less representative set of the overall variations of its class. Here, nothing can be said about which image is closest to y. For the datasets considered, this case has not occurred.

Considering all cases, it can be seen that the proposed set estimation method can generate test images perfectly in at least 90% of the cases and generated approximations in around 9% cases.

7. Comparison of classification accuracies and dataset description

In the earlier subsections, new face images were generated in each face class using algorithm 1 and new face images were generated for some classes using algorithm 2. In order to maintain uniformity, we took the new face images of algorithm 1 only, since four images (refer to remark 4 of Section 5.2.1) are generated for every class in this method. The experiment is done to check the performance of the classification technique using the newly generated images in the training set and keeping the test set as constant. Varying the training set consisting of the original face data along with the newly generated faces the comparisons are given in Table 2. The following procedure describes the technique used for this experiment for each face dataset and for each dimensionality reduction scheme.

(1) Perform dimensionality reduction by taking into consideration all the images in the training set of face database.
(2) Consider a training set consisting of five images from each face class. The test set consists of the remaining images in the database.
(3) For each face class, obtain the corresponding 5 transformed image vectors using the reduced features.
(4) Classify every image in the test set with the help of training set by using nearest neighbor classifier and obtain the correct classification rate in percentage.
(5) Obtain MST of the five transformed image vectors for each face class.
(6) Using algorithm 1, generate four new images for each face class. In the second training set, for each face class, add the four generated images to make the number of training images as 9. The new training set may be named as training set 2.
(7) Classify every image in the test set with the help of training set 2 by using nearest neighbor classifier, and obtain the correct classification rate in percentage.

Person 1 Intermediate images Person 2

Fig. 6. Image set 4. In each row, the smooth change of faces of intersecting classes is shown. Person 1, Intermediate images and Person 2.

Fig. 7. Resultant set 5. For each row, the image in the third column is the generated distorted face from the first two images, where the corresponding face classes are non-intersecting.

The FIA expression dataset used for testing in the context of the proposed face generation scheme consists of 13 classes having 75 images in each class. 5 face images are considered in the first training set for each face class. Test set therefore consists of 910 images. For AR dataset 40 classes of images are considered with 13 images in each class. Five images are considered in the first training set from each face class. For both the dataset four images are generated and added to every face class in the training set by Algorithm 1 and thus in the second training set the total number of images per class becomes 9. The results in Table 2 indicate that the recognition rates are much higher in the cases, where the generated face features are added from intra classes.

7.1. Test of significance

Table 3 indicates that the performance of proposed method is better when the synthesized images are generated by the methods proposed are added in the training set and used for recognition rate calculations. Z-test for the equality of proportions is employed to check the performance of the proposed method. The test statistic is given by

$$a = \frac{p_1 - p_2}{\sqrt{p(1-p)((1/n_1) + (1/n_2))}} \tag{7}$$

where $p_1 = \frac{x_1}{n_1}$, x_1 is the number of correctly classified images out of n_1 images using the PCA nearest neighbor classifier when only the original faces are considered. $p_2 = \frac{x_2}{n_2}$, where x_2 is the number of correctly classified images out of n_2 images using the PCA nearest neighbor classifier when the generated images are considered along with the original faces and

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

The observed values of a are greater than 3 for all the cases and the values lie outside not only the 95% confidence interval $(-1.96, 1.96)$ but also the 99% confidence interval $(-2.575, 2.575)$. Thus, we reject the hypothesis that the two proportions are same.

Another experiment is conducted to check whether the synthetic images alone can perform classification better than the original face images. For this purpose, different training sets are made with different number of synthesized images. The experiment is carried out on different number of images for training sets. The results on two such sizes are shown in Table 3. In the first such training set, for each edge in MST, exactly one synthesized image is generated (i.e., the corresponding λ value = 0.5), and thus the number of training images for each class is 4. In the second training set,

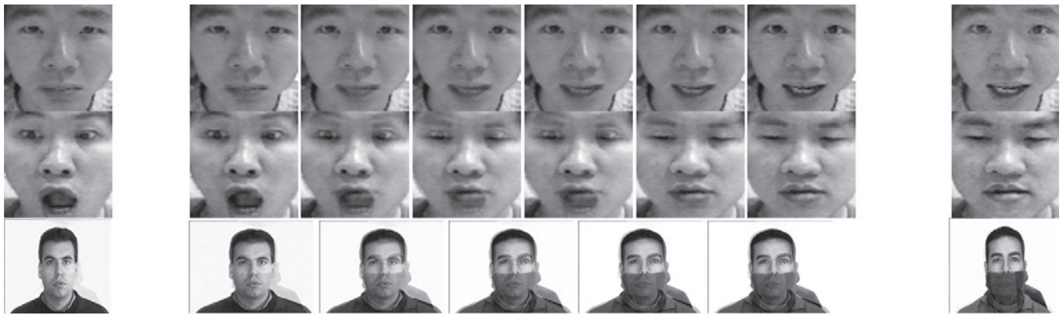

Fig. 8. Training image, test image and intermediary images for resultant set 6, where training and test images are from the same class. Face in training set (x), intermediate images generated varying r, generated test image (y).

Fig. 9. Generated face images with closest approximation of y using x.

for each edge in MST, four synthesized images are generated, where the corresponding λ values are 0.2, 0.4, 0.6 and 0.8. Thus the number of training images for each class is 16. This experiment is carried on all dimensionality reduction methods. The results on PCA and 2DPCA are shown in Table 3. In general, it has been observed that the recognition rates are improving as the size of the training set is increased for every dimensionality reduction method.

8. Conclusions

In this paper we put emphasis on face generation and face verification on the estimated face classes. The whole method is applied on 2D frontal faces. We have not considered any model for faces and we are not working in 3D space. The features are shared in two ways by using inter and intra face classes. The selection of meaningful faces from the generated faces is successfully done using set estimated disc radius. From the experiment it is shown that the distorted faces are automatically rejected in case of both FIA and AR datasets.

In the statistical approach mentioned in the present paper, the modeling of a face class would be more accurate if more points are present in the training set corresponding to that class. Many new images are generated in a face class with the help of training sample points from that class, resulting in face images with different expressions as well as different artifacts. The convex combinations are considered between those points only which form edge of MST. The smooth variation in face images from one class to another class is also shown.

Table 2
Result of Recognition rates.

Face database number of face classes, total number of training images	Method applied for dimensionality reduction	Recognition rate with the original face images in the database	Recognition rate with the synthetic faces added to the original training set
FIA	PCA	91	94
13 classes	KPCA	93	96
65 Training images in subspace	2DPCA	94	96
	PCA-LDA	88	90
AR	PCA	87	90
40 Classes	KPCA	93	94
200 Training images in subspace	2DPCA	92	94
	PCA-LDA	88	89

Table 3
Variation of recognition rates, where the training set consists of only the generated images.

Face database number of face classes, total number of training images	Method applied	λ Values	Number of images generated from each face class using the proposed method	Recognition rate with the original face images in the database five images per class	Recognition rate with the synthetic faces generated
FIA	PCA	0.5	4	91	90
		0.2, 0.4, 0.6, 0.8	16	93	96.5
AR	2DPCA	0.5	4	87	90.5
		0.2,0.4, 0.6,0.8	16	91	96

PSNR value is used to judge the quality of an image with respect to a given image. The PSNR values for all generated images are found to be in the interval of 25–40 dB. This result reflects good and acceptable performance of the proposed methods.

Another new aspect of the method is to classify the newly generated face into respective class. Nearest neighbor classifier is used for classifying the generated faces, and the classification is tallying with the intuition. As an extension, the whole process has also been applied and tested with other dimensionality reduction techniques like KPCA, 2DPCA, PCA-LDA and the classification results are acceptable. When the generated images alone are used in the training set, the resultant recognition rates are also found to be satisfactory. We have taken reasonably low number of training images It is expected that the results will improve if the size of the training set is larger.

In this method we have tested the set estimation procedure using Euclidean distance, since the theory of set estimation is dependent on Euclidean distance. The method is performing well and the procedure is verified theoretically and experimentally. It needs to be investigated theoretically if the other distance measures used. The generalization capability of generating the test images using the set estimation method is shown for images which are not present in the training dataset.

References

[1] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, Comput. Vis. Image Understand. 101 (1) (2004) 1–15.
[2] H. Tao, R. Lopez, T. Huang, Tracking facial features using probabilistic network, in: Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98), 1998, pp. 166.
[3] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1063–1074.
[4] T.F. Cootes, K. Walker, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685.
[5] L. Huang, C. Su, Facial expression synthesis using manifold learning and brief propagation, Soft Comput. (10) (2006) 1193–1200.
[6] Z. Liu, Y. Shan, Z. Zhang, Expressive Expression Mapping with Ratio Images, in: ACM Siggraph, vol. 10, 2001, pp. 271–276.
[7] T. Strothotte, S. Schlechtweg, Non-Photorealistic Computer Graphics, Modeling, Rendering, and Animation, Morgan Kaufmann, San Francisco, 2002.
[8] Q. Zhang, Z. Liu, B. Guo, D. Terzopoulos, H.Y. Shum', Geometry driven photorealistic facial expression synthesis, IEEE Trans. Visual. Comput. Graph. 12 (1) (2006) 48–60.
[9] Y. Liu, K. Schmidt, J. Cohn, S. Mitra, Facial asymmetry quantification for expression-invariant human identification, CVIU 91 (2003) 138–159.
[10] L. Wecker, F. Samavati, M. Gavrilova, A multiresolution approach to iris synthesis, Comput. Graph. 34 (3) (2010) 468–478.
[11] I. Cohen, N. Sebe, A. Garg, L. Chen, T. Huang, Facial expression recognition from video sequences: temporal and static modeling, CVIU 91 (1–2) (2003) 160–187.
[12] H. Pyun, Y. Kim, W. Chae, H.W. Kang, S.Y. Shin, An example-based approach for facial expression cloning, in: Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer, Animation, 2003.
[13] F.H. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D. Salesin, Synthesizing realistic facial expressions from photographs, SIGGRAPH (1998) 75–84.
[14] T. Beier, S. Neely, Feature-based image metamorphosis, Comput. Graph. 26 (2) (1992) 35–42.
[15] C.A. Murthy, On Consistent Estimation of Classes in R^2 in the Context of Cluster Analysis, Ph.D thesis, Indian Statistical Institute, Calcutta India, 1988.
[16] M. Datta, C.A. Murthy, Class specific threshold selection in face space using set estimation technique for RGB color components, In: International Conference on Pattern Recognition and Machine Intelligence (PReMI' 09), 2009, pp. 351–356.
[17] U. Grenander, Abstract Inference, John Wiley, New York, 1981.
[18] H. Edelsbrunner, D.G. Kirkpatrickand, R. Seidel, On the Shape of a set of points in a plane, IEEE Trans. Inform. Theory (IT-29) (1983) 551–559.
[19] D.P. Mandal, C.A. Murthy, S.K. Pal, Determining the shape of a pattern class from sampled points: extension to R^N, Int. J. Gen. Syst. 26 (4) (1997) 293–320.
[20] D.P. Mandal, C.A. Murthy, Selection of alpha for alpha-hull and formulation of fuzzy alpha-hull in R^2, Int. J. Uncertain. Fuzz. Knowl. Based Syst. 3 (4) (1995) 401–417.
[21] A.M. Martinez, R. Benavente, The AR Face Database, CVC Technical, Report #24, 1998.
[22] Face In Action (FIA) Face Video Database, AMP, CMU. <http://amp.ece.cmu.edu/projects/FIADataCollection/>.
[23] J. Ruiz-del-Solar, P. Navarrete, Eigenspace-based face recognition: a comparative study of different approaches, IEEE Trans. Syst. Man Cybern., Part C 35 (3) (2005) 315–325.
[24] D. Maio, D. Maltoni, Real-time face location on gray-scale static images, Pattern Recognit. 33 (9) (2005) 1525–1539.