3D Age Progression Prediction in Children’s Faces with a Small Exemplar-Image Set

CHENG-TA SHEN1, FAY HUANG2, WAN-HUA LU3, SHENG-WEN SHIH1 AND HONG-YUAN MARK LIAO3

1Department of Computer Science and Information Engineering
National Chi Nan University
Nantou, 545 Taiwan

2Institute of Computer Science and Information Engineering
National Ilan University
Ilan, 360 Taiwan

3Institute of Information Science
Academia Sinica
Taipei, 115 Taiwan

This work aims to develop a system for predicting age progression in children’s faces from a small exemplar-image set, which is a critical task to assist in the search for missing children. The proposed method consists of a facial component extraction module, a facial component distance measurement module, and a face synthesis module. It is developed based on the assumption that two similar facial components of two children will retain similar when they grow up. Two different distance measures, namely the learning-based Mahalanobis distance and the curvature-weighted plus bending-energy distance, are employed to select similar facial components from an aging database. The growth curve of each facial component is used to predict the shape, size, and location of each component at a different age. The thin plate spline method is applied to synthesize a 3D face model from the predicted components by minimizing the bending energy. Experiments are conducted to test the proposed method with various subjects and the results show that the proposed method yields promising results.

Keywords: age progression prediction, growth curve, face image synthesis, missing children search, metric learning

1. INTRODUCTION

For the movie, “The Curious Case of Benjamin Button,” which described a man who ages in reverse, the face age progression prediction (FAPP) technique was used to develop the actor’s rejuvenating effects from the predicted appearances. But predicting face progression has become an important topic with many practical applications. FAPP can be used for many other applications, such as cosmetology, biometrics, forensic art, and searching for missing children. It should be noted that FAPP differs from aging estimation, wherein an input face image is used to estimate the individual’s age, e.g., [1-4]. Therefore, aging estimation can be regarded as the inverse problem of FAPP.

In this work, we study the FAPP problem technique applied for missing children searching. According to a report from the U.S. Department of Justice, there are over 790,000 children (younger than 18) reported missing each year [5]. Since many of the

Received April 4, 2012; accepted June 5, 2012.
Communicated by Chung-Lin Huang.
* This work was supported in part by the National Science Council, Taiwan, under Grant No. NSC 98-2221-E-260-022-MY3 and by Ministry of Economic Affairs, Taiwan, under Grant No. 100-EC-17-A-02-S1-032.
missing children were lost/kidnapped when they were very young, as the missing children grow up their faces may appear quite differently from the photos provided by their parents. Therefore, if a young missing child cannot be found soon, facial change often make it difficult to find the child. If we can predict a child’s facial components according to the growth curves of the facial components of his/her parents, siblings, or some similar people, this may increase the chance of a successful search. The main difficulties of FAPP are as follows.

• **Growth rate variation:** Different people have different rates of aging and people’s appearance during their growth is influenced by many factors, such as race, heredity [2], social factors [6], health, lifestyle, working conditions, etc.
• **Gender:** Males and females have different growth phases, and a female’s growth phase usually occurs earlier than that of a male.
• **Aging catalyst:** Some catalysts lead to premature aging in humans, including long exposure to the sun, smoking, susceptibility to disease, pressure and mood of life, illness, abrupt weight change, etc.
• **Data collection:** Establishing the growth model for different persons requires a huge amount of data, and collecting such a large database with so many subjects with images at different ages is very difficult. At present, the number of subjects in the publicly available dataset (e.g., [7]) is less than 100 persons which is far from sufficient for estimating a reliable face aging model.

These issues make FAPP a very challenging problem. In this work, a component-exemplar-based FAPP method is proposed. The proposed method is different from many existing methods, such as [8-11], which all depend on having complete face images. In contrast, the prediction of each facial component is computed individually with our method. Given an image of a missing child, we first extract the facial components, such as the eyes, the nose, the mouth and the face shape of the child. For each facial component of the child, similar facial components of different children in a training database are extracted. Under the assumption that facial components of two similar children will remain similar when they grow up, we can use their growth curves to predict the facial
components of the missing child at different ages. The predicted facial components are then merged to a complete face in frontal view using the thin plate spline (TPS) method.

The remainder of this paper is organized as follows. Section 2 describes related work. Section 3 introduces the proposed method. Section 4 shows the experimental results of the proposed method. Conclusions and future work are described in Section 5.

2. RELATED WORKS

As human faces change with age, facial aging effects are attributed mainly to the growth of craniofacial and skin related deformations associated with the introduction of wrinkles and reduction of muscle strength [12, 13]. In particular, the craniofacial growth (human face shape) is the most obvious from birth to adulthood. Including the growth of the face size, eyes, nose, mouth and skull, and the augmentation of the chin and cheek, each type of growth affects the facial appearance. Although there are many changes in facial skin over time, the skin change is more subtle in proportion to the craniofacial growth, such as pores thickening, skin color changes, and even mustache growth.

Existing FAPP methods can be roughly classified into anthropometry-model-based approaches, AAM-based approaches and landmark-based approaches. Detailed surveys of existing FAPP methods can be found in [14-16]. In general, anthropometry-model-based methods estimate facial aging with a face model controlled by a high-dimensional parameter vector. Lin et al. adopted anthropometric data from [17] to create a statistical growth function that was used to simulate the aging process of 3D graphic models for a Caucasian boy, an African-American boy and an Asian boy, respectively [18]. Ramanathan and Chellappa [9] proposed a craniofacial growth model to predict the progression of the human face under the age of eighteen years old. Ariyarathne and Dharmaratne also proposed a very similar method for an age progression prediction [19]. They employed the anthropometric statistical data of 2,325 Caucasian people collected by Farkas [17] to predict the variations of facial landmarks. Then the variation of every pixel in the entire image was determined based on the variations of the landmarks with an interpolation method to nonlinearily warp the input image. Therefore, their method can only handle one input image. Furthermore, if there is a large age difference between the source image and the age-enhanced image, the interpolated nonlinear warping computed with facial landmarks will be inaccurate.

The AAM-based approaches are developed based on the active appearance models. Suo et al. proposed a multi-resolution AAM model which categorizes face images in the same age group by using a hierarchical and-or graph to account for the large variations of facial structures [20-22]. Suo et al. modeled the aging procedure as a dynamic Markov process to predict adult aging faces, which provides satisfactory results for predicting adult aging, but is not able to handle the FAPP problem for children [20]. Luu et al. integrated the AAM and the support vector regression (SVR) to learn the aging function, using a training database to construct standard feature vectors at different ages [23]. The difference between siblings’ feature vector and the standard feature vector are used to adjust the input feature vector. The resulting feature vector is converted into a photo with the inverse AAM. However, their method can be applied with only one sibling photo, and if the sibling does not look like the subject the prediction result is unreliable. Lanitis
et al. employed the AAM technique to build a statistical face model to extract changing components of facial shapes and intensity from training data [8]. Then the facial parameters in the proposed aging function were optimized by using the genetic algorithm (GA). The proposed aging function can not only estimate the age of the child facial image but also reconstruct the appearance of the human face at any age. Geng et al. also applied the AAM for reconstructing human faces and estimating ages [10]. They defined an aging pattern as a sequence of personal face images sorted in time order and construct a subspace by using principle component analysis (PCA), and the constructed subspace captured the main aging variation in the data set. Since the aging pattern vector is incomplete, they used an EM-like algorithm to learn a representative subspace. Then, an unseen input face image was projected into the subspace and an optimal aging pattern minimizing the error between the input image and the reconstructed image was chosen to represent the aging pattern of this input face image.

Fig. 2. 3D face model; (a) A 2-D face image; (b) The converted 3D face model without adjustment; (c) The converted 3D face model after adjustment to have a neutral expression.

The landmark-based approaches construct the growth function of the face shape defined with a set of facial landmarks. The facial landmarks can be determined either manually [18, 24-26] or automatically [27]. Hill et al. used the PCA technique to generate a shape model and a texture model, and the proposed method could compensate for the face orientation and expression to achieve reliable face aging predictions [25]. Liang et al. improved the method of [25] by introducing a procrustes method to align facial features. So, the artifacts in the face image composed using the PCA bases can be minimized [26]. Gandhi applied the SVR technique to learn an aging function and used the aging function and an existing texture transformation method to perform face aging progression prediction, though that study did not test the FAPP result for children [24]. Scherbaum et al. reconstructed 3D face shapes from images and the 3D vertices and texture images were used to construct aging functions with SVR. The aging transformations of the shape and the texture were then applied to synthesize images of the age-enhanced faces [27].

These three categories of methods all require a large training set containing data/images of many different peoples. They all try to find the mean growth model of the majority to perform FAPP with statistical anthropometric data, PCA, AAM or SVR. Therefore, the main drawback of those three kinds of methods is that they tend to ignore individual differences. Although the best way to approach the FAPP of a missing child is to use the photos of his/her immediate relatives at different ages, but the number of immediate relatives may be very limited, so existing methods may have limited application.
To find the growth model for the whole face of an input photo from a small training-image set is difficult, unless the image set contains only a series of photos of a very similar-looking person, e.g., an identical twin. Conversely, because immediate relatives share common inheritance, it is easier to find almost identical facial components than to find almost identical faces among relatives. Therefore, the main contribution of this work is to propose a facial-component-based FAPP method that is suitable to perform FAPP with a very small training database.

Fig. 3. Facial component extraction; (a) Facial components of a subject; (b) Parameters of a facial component (take the eye component as an example).

3. PROPOSED METHOD FOR PREDICTING AGE PROGRESSION

A flowchart for the proposed algorithm is shown in Fig. 1. Suppose that there is an aging database containing face images of $I$ subjects at different ages, and the aging database does not have to include all-age images for each subject. Suppose that each image in the database has been converted into a 3D face model using a commercialized software package such as FaceGen [28]. Furthermore, to simplify the subsequent computation, all the generated 3D faces have been adjusted to have neutral expressions, as shown in Fig. 2.

To predict a progression for child’s appearance, a series of age-progressed images of the child are required. The input images are also converted into expression-neutralized 3D face meshes to construct a 3D aging database. The main idea of our method is to find subjects in the aging database who have similar facial components with the child and, then, to use the growth curve of the selected subjects to predict the appearance of the child at a different age, say age $\tilde{a}$. Ideally, in searching for subjects with similar facial components, 3D data converted from images taken at the same age should be used. However, with a small-scale aging database, it will be difficult to find facial components of subjects obtained exactly at the same age. Therefore, we have to categorize all age-progressed images into several age groups. According to the data provided by Farkas, different facial components, such as the width of the face, mouth and eye, have different growth spurt periods. Farkas categorized children’s age progression into two growth periods, i.e., early growth (1-6 years old) and late growth (6-16 years old) [17]. However,
since the children’s appearance changes very fast, it is not reasonable to compare the similarity of two facial components at ages with such a big difference. Hence, the age-progressed component sequence in this work is split into five age groups denoted as $A = \{A_1, A_2, A_3, A_4, A_5\}$, where $A_1$, $A_2$, $A_3$, $A_4$, and $A_5$ represent age groups of 0-3, 3-6, 6-11, 11-16, and 16-18 years old, respectively. It is assumed that the input images were taken when the missing child was in age groups $A' \subseteq A$.

3.1 Facial Components Extraction

The facial components adopted in this work include four 3D meshes representing the face shape, the mouth, the nose, and the eyes, respectively. It is assumed that the face is bilaterally symmetrical, so only one 3D mesh is used to represent the eye component. The 3D mesh of the other eye can be generated using a mirror operation. Let $g_{c,a,i} = \{g_{f,a,i}, g_{e,a,i}, g_{n,a,i}, g_{m,a,i}\}$ denote the facial component set of subject $i$ at age group $a \in A$, where subscripts $f$, $e$, $n$, and $m$ indicate face, eye, nose, and mouth, respectively (refer to Fig. 3 (a)). For convenience, let $\mathcal{C} = \{f, e, n, m\}$ and let the missing child be indexed by zero (i.e., the 0th subject). Also, let $h_{c,a,i}$ and $w_{c,a,i}$ denote the height and width of facial component $c (c \in \mathcal{C})$, respectively. Then, each facial component is defined as $g_{c,a,i} = \{l_{c,a,i}, \psi_{c,a,i}, M_{c,a,i}\}$ (refer to Fig. 3 (b)), where

- $l_{c,a,i} = \begin{bmatrix} x_{c,a,i} \\ y_{c,a,i} \\ w_{c,a,i} \end{bmatrix}$ is the normalized location of component $c$ and $[x_{c,a,i}, y_{c,a,i}]$ is the translation vector of the local coordinate system $\{x_c, y_c, O_c\}$ of component $c$ to the face coordinate system $\{x_f, y_f, O_f\}$;
- $\psi_{c,a,i} = \begin{bmatrix} \psi_{c,a,i} \\ \psi_{c,a,i} \\ \psi_{c,a,i} \end{bmatrix}$ is the relative size of component $c$ with respect to the face width;
- $M_{c,a,i}$ is the 3D mesh of component $c$ which is composed of a constant number of vertices $N_c$.

3.2 Distance between Facial Components

Given the constructed 3D aging database, we will describe how to select $K$ nearest neighbors to synthesize a predicted new face. In order to do that, we have to determine the similarity between two facial components. There are two aspects to be considered in determining the similarity between two facial components. We not only have to estimate the similarity between the 3D meshes of two facial components, but also have to estimate the similarity of their growth curves. The growth curve of a facial component includes the trajectories of the normalized location, i.e., $l_{c,a,i}$ and the relative size, i.e., $\psi_{c,a,i}$, as functions of $a$.

Since the Euclidean distance can directly reflect the dissimilarity of the location and size, the dissimilarity of the growth curves $L_{c,j} = \{l_{c,a,i} | a \in A'\}$ and $\psi_{c,j} = \{\psi_{c,a,i} | a \in A'\}$, where $A' \subseteq A$, are defined as:

$$D(L_{c,i}, L_{c,j}) = \sum_{a \in A'} ||l_{c,a,i} - l_{c,a,j}||,$$  \hspace{1cm} (1)

$$D(\psi_{c,i}, \psi_{c,j}) = \sum_{a \in A'} ||\psi_{c,a,i} - \psi_{c,a,j}||.$$  \hspace{1cm} (2)
Selecting a proper distance measure is essential in determining similar facial components. To assess the similarity of the 3D meshes of two subject’s facial components in the same age group, the two 3D meshes should be registered to have identical mass centers and identical width, i.e., \( w_{c,i} \). The mesh registration can be accomplished easily because their mesh structures are identical. After registering the meshes, the 3D coordinates of the \( N_{c} \) vertices of each registered mesh are concatenated to form a vector denoted as \( v_{c,i} \in \mathbb{R}^{3N_{c}} \). Then, the distance can be represented as \( D(v_{c,i}, v_{c,j}) \), where \( D(\cdot, \cdot) \) is a distance function to be discussed later. Because the Euclidean distance is not consistent with the distance between two meshes as perceived by most of the people, two distance functions, namely, the curvature-weighted plus the bending energy distance and the learning-based Mahalanobis distance, are considered in this work.

- **Curvature-Weighted (CW) + Bending-Energy (BE) Distance**

Euclidean distance is not suitable to measure the dissimilarity between two 3D meshes because all the vertices are treated equally. However, when comparing two meshes, people usually focus on the difference between two high-curvature surface patches. Therefore, we propose to weight the difference between surface patches using the absolute Gaussian curvature value [29]. Let \( \kappa_{q,i} \) and \( \kappa_{q,j} \) denote the Gaussian curvatures of the \( q \)th vertex of mesh \( M_{c,i} \) and \( M_{c,j} \), respectively. A weighting factor of the \( q \)th vertices can be defined as:

\[
\omega_{q} = \frac{1}{Z} \max(\kappa_{q,i}, \kappa_{q,j})
\]

where \( Z = \sum_{q=1}^{c} \omega_{q} \) is a normalization term. A \( 3N_{c} \times 3N_{c} \) block diagonal matrix \( W \) can be constructed whose \( q \)th diagonal block is given by

\[
W_{q} = \omega_{q} I_{3}
\]

where \( I_{3} \) is the \( 3 \times 3 \) identity matrix. Therefore, the weighted distance can be defined as follows.

\[
D_{w}(v_{c,i}, v_{c,j}) = \| W(v_{c,i}, v_{c,j}) \|
\]

where \( c \in \{ e, n, m \} \). The curvature-weighted distance works well for computing the distances between the meshes of the eye, the nose, and the mouth. However, when comparing two craniofacials that do not include the regions of eye, nose and mouth, the remaining surface is very smooth and the curvature values are very low. Therefore, the curvature-weighted distance is not suitable to compare the dissimilarity of two face shapes and, hence, the bending energy [30] is used instead to compute the distance between two face shape components, \( M_{f,i} \) and \( M_{f,j} \), denoted by \( D_{b}(v_{f,i}, v_{f,j}) \). Furthermore, because the goal of this work is to synthesize a frontal view of the missing child, the contour of the face dominates the face shape in the face synthesized image. Therefore, we compute only the bending energy of the face contours as a measure of the dissimilarity of two face shapes.
Learning-based Mahalanobis (LM) Distance

Evaluating the similarity between two human facial components is a subjective problem. Consequently, it may be imprudent to directly increase the weights of those high-curvature vertices. A more appropriate approach is to employ the supervised distance metric learning algorithm [31] to learn a Mahalanobis distance function with a positive definite matrix $\mathbf{\Omega}$, as shown in following,

$$D_M(v_{c,a,i}, v_{c,a,j}) = (v_{c,a,i} - v_{c,a,j})^T \mathbf{A} (v_{c,a,i} - v_{c,a,j}).$$

A prerequisite for the metric learning algorithm is to create a set of labeled pairwise training data,

$$T = \{(\ell_{c,a,i}, (v_{c,a,i}, v_{c,a,j})) | a \in \mathcal{A}, c \in \mathcal{C}, 1 \leq i, j \leq I\}$$

where the manually determined label is given by

$$\ell_{c,a,i} = \begin{cases} 0 & \text{if } (v_{c,a,i}, v_{c,a,j}) \text{ are dissimilar} \\ 1 & \text{otherwise} \end{cases}$$

In addition, an upper bound $u_b$ and a lower bound $l_b$ of the desired distance function should be provided. The metric learning algorithm [31] will automatically adjust the weighting matrix $\mathbf{A}$ trying to fulfill the following soft constraints.

$$\begin{cases} D_u(v_{c,a,i}, v_{c,a,j}) > l_b & \text{if } \ell_{c,a,i} = 0 \\ D_u(v_{c,a,i}, v_{c,a,j}) > u_b & \text{if } \ell_{c,a,i} = 1 \end{cases}$$

3.3 Selection of Similar Facial Components

Recall that a facial component $c$ is described by three parameters $l_{c,a,0}$, $\psi_{c,a,0}$, and $M_{c,a,0}$. Since the subsequent method for processing all the parameters of all the facial components are all the same, for simplicity, we will use $p_{a,i}$ represent every parameter of a facial component $c \in \mathcal{C}$, i.e., $p_{a,i} \in \mathcal{G}_{a,i}$. By using the aforementioned distance functions, we can select the $K$ closest parameters of $p_{a,0}$ from the aging database of the same age group. Let $\mathcal{S}$ denote the subject indices of the $K$ closest components of $p_{a,0}$.

3.4 Facial Component Synthesis

The facial component $p_{a,0}$ predicting equation is given by

$$p_{a,0} = \sum_{a = A} \omega_{a,0} (\sum_{i = b} \omega_{a,i} p_{a,i})$$

where $\omega_{a,0}$ and $\omega_{a,i}$ are two kinds of unit-sum weighting factors that are proportional to $e^{-\lambda |a-b|}$ (where $\lambda$ is an empirically determined constant) and $e^{-D(p_{a,0}, p_{a,i})}$. The former weighting factor favors the prediction results made based on facial components of similar ages. The latter one favors the prediction results made by similar facial components.
By combining the predicted parameters computed using Eq. (10), we have each component of the facial components, \( g_{c,a,i} = \{ l_{c,a,i}, \varphi_{c,a,i}, M_{c,a,i} \} \) for \( c \in C \).

The following example explains how Eq. (10) works (refer to Fig. 4). For clarity of illustration, we use face photos and specific ages rather than parameters of facial components and age groups in this example. Suppose that two photos of a missing child are provided which were taken when the child was 3 and 10 years old and the goal is to predict the appearance when the child is 16. From the aging database, we can find \( K \) most similar three-year-old children, denoted as \( S_3 \), and another \( K \) most similar ten-year-old children, denoted as \( S_{10} \), using the input photos. The 16-year-old photos of these two sets of children can be retrieved, but from the aging database, the 16-year-old photos of the children in \( S_3 \) can be used to compose a face, denoted as \( p_{3,16} \), with weight \( \omega_{3,16} \). The weight factor is inversely proportional to the distance between the face of the missing child and a subject in \( S_3 \). Likewise, another face, denoted as \( p_{10,16} \) can be composed using the 16-year-old photos of \( S_{10} \). Finally, \( p_{3,16} \) and \( p_{10,16} \) are used to predict the 16-year-old face of the child with weights \( \omega_{3,16} \) and \( \omega_{10,16} \), respectively. The weight of \( p_{10,16} \) is greater than that of \( p_{3,16} \) to favor the prediction results of closer ages.

Fig. 4. Example illustrating how to synthesize facial components.

![Fig. 4. Example illustrating how to synthesize facial components.](image)

Fig. 5. (a) Query input. Facial components selected with different distance measures, (b) the CW+BE distance and (c) the LM distance. The blue (red) rectangle indicates the two most similar (dissimilar) facial components extracted from the C-25 database.

### 3.5 Composition of a Face Image Using the Synthesized Components

When synthesizing an age-enhanced face, a face deformation problem needs to be
resolved. This problem arose because the locations and sizes of the eye, nose and mouth have been changed, making them inconsistent to the face mesh. Therefore, the face mesh has to be adjusted according to the new locations and sizes of the facial components using the thin plate spline (TPS) method [30]. The contour points of the eye, nose, mouth and face are used as the control points while the other vertices of the face mesh are adjusted. Finally, the texture image of the child at age \( \max(A') \) is mapped onto the 3D mesh to generate a frontal face image.

4. EXPERIMENTAL RESULTS

To evaluate different FAPP methods, we use the FG-NET aging database [7] to predict the face age progression of Caucasian people. The database is a publicly available image database including 1,002 face images of 82 subjects at different ages. Out of the 82 subjects, only subjects having at least one photo in each age group are selected to construct our aging database. Therefore, the aging database contains 125 photos of 25 subjects, which is a small database of Caucasian people and will be referred to as the C-25 database hereafter. We also use a very small aging database of the Jackson Five to predict the face age progression of Michael Jackson with the photos of his three siblings.

For training the LM distance, we use FaceGen to manually synthesize 100 similar pairs and another 100 dissimilar pairs of a facial component. For example, we can manually synthesize a dissimilar nose pair containing a snub nose and a hooked nose. Notably, because the training data are synthesized, the metric learning is totally independent of the subsequent 3D face aging progression prediction. Also, in the following experiments when we try to predict the age-enhanced face of a person in the C-25 database, the data of that person will be excluded from the prediction process.

The first experiment is to verify the effectiveness of both the facial component selection procedure and the face synthesis procedure our method. Figure 5 shows the results of facial component selection with different distance measures. The two most similar and the two most dissimilar components are enclosed in blue and red rectangles, respectively. For the case of mouth component, the dissimilar ones have thicker lips than the most similar ones. The similar face shapes are indeed more similar to the input face shape than the dissimilar ones. The results of both the CW+BE distance and the LM distance are satisfactory. However, since the LM distance has the ability to learn the difference between shapes perceived by people, given sufficient training data, the LM distance may outperform the CW+BE distance.

Before we show that the proposed method can be used to predict the appearance of a child at a different age, it has to be verified that the face synthesis procedure can reproduce a face similar to the original one of the same age (i.e., \( \hat{a} \in A' \)) using the selected similar facial components. The faces synthesized using the most similar and the most dissimilar \( K \) facial components/parameters are shown in Table 1. The parameter \( K \) is set to three throughout all the experiments. The results show that faces synthesized using similar components are similar to the original ones whereas faces synthesized using the most dissimilar components are quite different from the original ones.

In the second experiment, the goal is to test the performance of the proposed method in predicting the progression appearance of a child, i.e., \( \hat{a} \notin A' \), with a single input
image so as to compare the results with those computed with the craniofacial growth (CG) model [9], which will be referred to as the CG method hereafter. For convenience, the proposed methods with the CW+BE distance measure and the LM distance measure are referred to as the CW+BE method and the LM method, respectively. Table 2 shows the prediction results where the first column is the input image, the second column is the frontal view of the reconstructed 3D face, and the last column shows the actual photo at the specified age. The faces predicted with the CG method, the CW+BE method and the LM method are shown on columns two, four, and five, respectively. The results show that the proposed method is comparable to the CG method. However, if the person in the input photo appears to have baby fat (see the seventh row in Table 2), then the proposed method can achieve a better result.

Table 1. Results of the proposed face synthesis procedure.

<table>
<thead>
<tr>
<th>Input image</th>
<th>Reconstructed face</th>
<th>CW+BE distance</th>
<th>LM distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similar</td>
<td>Dissimilar</td>
<td>Similar</td>
<td>Dissimilar</td>
</tr>
</tbody>
</table>

A subjective evaluation is conducted to assess the performance of different methods. Instead of directly asking experiment participants to subjectively assign a score for each prediction result, we invited 10 responsible subjects to convert all the ground truth photos and all the CG prediction results to 3D face models using FaceGen. In the conversion process, each subject can try different parameters of the software program to make the 3D face look like the corresponding 2-D photo. The degree of freedom of FaceGen is large enough for 3D face reconstruction. The reconstructed 3D models of the ground truth and the CG predicted photos are shown in Table 3 which shows that the reconstruction process is satisfactory. The benefit of converting all 2D photos into 3D face models is that we can directly compute the error between a predicted 3D face model and its corresponding ground truth 3D face model using the bending energy technique [30]. The box plot of the bending energy errors of the CW+BE, LM, and CG methods are shown in Fig. 6. Notably, the bending energy of the CG prediction results of the person on the sixth row of Table 2 shows the greatest variation. This is because the face orientations the CG predicted and the ground truth are not the same, which increases the variation of the converted 3D models. The subjective evaluation results show that the proposed LM and CW+BE methods are slightly better than the CG method.

Since the proposed method is able to accept multiple input images at different ages, the third experiment is to test the proposed method with multiple input images. Tables 4 and 5 show the prediction results computed using two and three input images of the same subject at different ages, respectively. Fig. 7 shows the subjective evaluations of the predicting results with one to three input images. The results show that multiple input images can provide better prediction performance.
The last experiment is to verify the proposed method with a very small database composed of 15 aging photos of Michael Jackson’s three siblings, referred to as the MJ-3 database. The photos of his siblings at five different ages can be easily retrieved from the Internet. The prediction results are shown in Table 6. The prediction results of the CG method are shown on the second column of Table 6. Although the CG method was developed based on anthropometric data of Caucasian people, the prediction result shown on the first row is still satisfactory. However, since the CG method merely computes a nonlinear deformation of the 2D photo, the prediction result will have a childish face if the age difference between the input image and the age-enhanced image is too large (see the second and the third rows of the second column of Table 6). But, because the proposed methods use the growth characteristics of his siblings, the prediction results do not have the childish artifact. Moreover, since the MJ-3 database contains only three persons, the three most similar facial components selected by both the CW+BE method and the LM method are the same. Therefore, the prediction results shown on the third and the fourth columns are very similar. What makes the prediction results of the two methods slightly different is the weights of the three exemplars, which are determined by different distance measures.

### Table 2. Prediction results with a single input image.

<table>
<thead>
<tr>
<th>Input image</th>
<th>CG method</th>
<th>Reconstructed 3-D face</th>
<th>CW+BE method</th>
<th>LM method</th>
<th>Ground truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 yrs</td>
<td>5 yrs</td>
<td>4-6 yrs</td>
<td>4-6 yrs</td>
<td>5 yrs</td>
<td></td>
</tr>
<tr>
<td>9 yrs</td>
<td>12 yrs</td>
<td>12-15 yrs</td>
<td>12-15 yrs</td>
<td>12 yrs</td>
<td></td>
</tr>
<tr>
<td>8 yrs</td>
<td>12 yrs</td>
<td>12-15 yrs</td>
<td>12-15 yrs</td>
<td>12 yrs</td>
<td></td>
</tr>
<tr>
<td>6 yrs</td>
<td>12 yrs</td>
<td>12-15 yrs</td>
<td>12-15 yrs</td>
<td>12 yrs</td>
<td></td>
</tr>
<tr>
<td>11 yrs</td>
<td>18 yrs</td>
<td>16-18 yrs</td>
<td>16-18 yrs</td>
<td>18 yrs</td>
<td></td>
</tr>
<tr>
<td>8 yrs</td>
<td>16 yrs</td>
<td>16-18 yrs</td>
<td>16-18 yrs</td>
<td>16 yrs</td>
<td></td>
</tr>
<tr>
<td>10 yrs</td>
<td>18 yrs</td>
<td>16-18 yrs</td>
<td>16-18 yrs</td>
<td>18 yrs</td>
<td></td>
</tr>
<tr>
<td>7 yrs</td>
<td>16 yrs</td>
<td>16-18 yrs</td>
<td>16-18 yrs</td>
<td>16 yrs</td>
<td></td>
</tr>
</tbody>
</table>
The results of the last experiment show that when the age difference between the input image and the age-enhanced image is small, the amount of the nonlinear deformation computed with the CG method is small and the prediction results will be good. In that case, the proposed methods are comparable to the CG method. However, when the age difference is large, the CG method will suffer from the childish artifact. Then the proposed methods will outperform the CG method.

Table 3. Three sets of 3D face models reconstructed from the CG prediction results and the ground truth (GT) photos by 10 different subjects.

<table>
<thead>
<tr>
<th>GT</th>
<th>CG</th>
<th>GT</th>
<th>CG</th>
<th>GT</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
</tbody>
</table>

The average computation time for the CG method is about 0.94 seconds whereas the computation time for the LM and the CW+BE methods with one input image is 0.8 and 1.8 seconds, respectively. With three input images, the LM method and the CW+BE method cost 1.4 and 2.8 seconds, respectively. When the size of the database is large or when more input images are available, the computation time will ordinarily increase. However, since FAPP is not a real time application, the computation time of the proposed method suffices most FAPP applications.
5. CONCLUSION

FAPP techniques can be used for many applications, such as cosmetology, biometrics, forensic art, and missing children searching. Existing methods require a large database to find the mean growth model for FAPP, but the mean growth model tends to ignore individual differences. Therefore, the best way to approach the FAPP of a missing child is to use the photos of his/her immediate relatives in different ages. However, there may be only limited photos of immediate relatives, and so existing methods may not be applicable. While it is difficult to find the growth model of the whole face from a small familial database, it is easy to find almost identical facial components from the familial database because immediate relatives share a common inheritance. The main contribution of this work is to propose a facial component based FAPP method. Given an image of a missing child, we first extract facial components, such as the eyes, the nose, the mouth and the face shape of the child. For each facial component of the child, similar facial components of different children in a training database are extracted, and their growth curves are used to predict the facial components of the missing child. The similarity of two different facial components is evaluated with two different distance measures, i.e., the curvature-weighted plus bending-energy distance, and the learning-based Mahalanobis distance. The experimental results show that the face images synthesized with our method are similar to the ground truth photos.

Fig. 6. The bending energy error of the CG (blue), CW+BE (red), and LM (black) methods.

Fig. 7. Subjective evaluation results of the CW+BE and LM methods with single (blue), two (green) and three (red) input images, respectively.
### Table 4. Prediction results with two input images.

<table>
<thead>
<tr>
<th>Input image</th>
<th>CW+BE method</th>
<th>LM method</th>
<th>Ground truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 yrs</td>
<td>16 yrs</td>
<td>16-18 yrs</td>
<td>18 yrs</td>
</tr>
<tr>
<td>5 yrs</td>
<td>9 yrs</td>
<td>12-15 yrs</td>
<td>12 yrs</td>
</tr>
<tr>
<td>8 yrs</td>
<td>12 yrs</td>
<td>16-18 yrs</td>
<td>16 yrs</td>
</tr>
</tbody>
</table>

### Table 5. Prediction results with three input images.

<table>
<thead>
<tr>
<th>Input image</th>
<th>CW+BE method</th>
<th>LM method</th>
<th>Ground truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 yrs</td>
<td>10 yrs</td>
<td>16 yrs</td>
<td>16-18 yrs</td>
</tr>
<tr>
<td>3 yrs</td>
<td>5 yrs</td>
<td>12-15 yrs</td>
<td>12 yrs</td>
</tr>
<tr>
<td>5 yrs</td>
<td>8 yrs</td>
<td>12 yrs</td>
<td>16-18 yrs</td>
</tr>
</tbody>
</table>

### Table 6. Michael Jackson.

<table>
<thead>
<tr>
<th>Input image</th>
<th>CG method</th>
<th>CW+BE method</th>
<th>LM method</th>
<th>Ground truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 yrs</td>
<td>18 yrs</td>
<td>16-18 yrs</td>
<td>18 yrs</td>
<td></td>
</tr>
<tr>
<td>6 yrs</td>
<td>13 yrs</td>
<td>12-15 yrs</td>
<td>13 yrs</td>
<td></td>
</tr>
<tr>
<td>6 yrs</td>
<td>18 yrs</td>
<td>16-18 yrs</td>
<td>18 yrs</td>
<td></td>
</tr>
</tbody>
</table>

### REFERENCES


Cheng-Ta Shen (沈政達) received the M.S. degree in 2007 and is currently pursuing the Ph.D. degree in the Department of Computer Science and Information Engineering, both from the National Chi Nan University, Taiwan. His current research interests include computer vision, human-computer interaction and image processing.

Fay Huang (黃于飛) received the M.Sc. degree with honour (second class, first division) in Pure Mathematics and the Ph.D. degree in Computer Science (2002) from The University of Auckland, New Zealand. She worked as a Postdoctoral Fellow at Institute of Information Science, Academic Sinica, from January 2003 to July 2004. She is currently an Assistant Professor at Institute of Computer Science and Information Engineering, National Ilan University, Taiwan. Fay’s research interests include computer vision and computer graphics, among which she is particularly interested in applications related to virtual reality and computer art.
Wan-Hua Lu (呂婉華) received the M.S. degree in Computer Science and Information Engineering from the National Chi Nan University, Taiwan, in 2010. She is currently an Engineer with HTC Corp., Taipei, Taiwan. Her research interests include computer vision and image processing.

Sheng-Wen Shih (石勝文) received the M.S. and Ph.D. degrees in Electrical Engineering from the National Taiwan University, Taiwan, in 1990 and 1996, respectively. From January to July 1997, he was a Postdoctoral Associate with the Image Formation and Processing Laboratory, Beckman Institute, University of Illinois, Urbana-Champaign. He is currently a Professor with the Department of Computer Science and Information Engineering, National Chi Nan University, Taiwan. His research interests include computer vision, biometrics, and human–computer interaction.

Hong-Yuan Mark Liao (廖弘源) received the M.S. and Ph.D. degrees in Electrical Engineering from Northwestern University, Evanston, IL, in 1985 and 1990, respectively. He is jointly appointed as a Professor of the Department of Computer Science and Information Engineering, National Chiao Tung University and the Department of Electrical Engineering, National Cheng Kung University. From 2009 to 2012, he was jointly appointed as the Multimedia Information Chair Professor of National Chung Hsing University. Since August 2010, he has been appointed as Adjunct Chair Professor of Chung Yuan Christian University. He served as the conference co-chair of 2004 ICME and technical co-chair of 2007 ICME. He is on the editorial boards of the IEEE Transactions on Image Processing and the IEEE Signal Processing Magazine. He was an Associate Editor of IEEE Transactions on Multimedia during 1998-2001, and the IEEE Transactions on Information Forensics and Security during 2009-2012. His current research interests include multimedia signal processing, video-based surveillance systems, content-based multimedia retrieval, and multimedia protection.