A Novel Technique for Sketch to Photo Synthesis

Pulak Purkait*
ECSU, Indian Statistical Institute
203, B. T. Road, Kolkata - 108
West Bengal, India
pulak_r@isical.ac.in

Bhabatosh Chanda
ECSU, Indian Statistical Institute
203, B. T. Road, Kolkata - 108
West Bengal, India
chanda@isical.ac.in

Shrikant Kulkarni
Dept. of ECE, National Institute of Technology
Karnataka, Surathkal
Karnataka, India
shrikant06@ieee.org

ABSTRACT

We propose a novel pseudo photo generation system from a sketch image. All the training images and the input test sketch are warped around its mean shape to get shape free images. Based on local geometry preserving manifold learning method called locally linear embedding (LLE), a shape free pseudo-photo is synthesized by block matching. Then the photo of actual shape is constructed with the help of an Active Shape Model(ASM) and a neural network built on shape control point coordinate pair for training sketch-photo samples. We experimented over 300 sketch-photo image pairs and output synthesized photo for every test image is encouraging.

1. INTRODUCTION

Researchers today are actively making smart environments. These environments such as rooms, cars, offices and stores are made by smart visual, audio and touch sensitive applications. The next generation environment would give machines perceptual ability that allow them to interact people with their speech, gesture, and even body languages. Most of the next generation equipment will be touch sensitive and simple Sketch-Based Interfaces. A growing number of applications on those equipments have started to use simple sketches as a first step towards interpretation human action, intension and behavior. So face photo synthesis from a simple sketch can be an important component of such system.

One of the important application of face recognition is to automatic retrieval of photo from database for a query image. It can help the crime branch to narrow down the suspect quickly. Unfortunately in most of the cases, photo of the suspect is not available. To deal with such problem sketch drawing by some artist with cooperation from an eyewitness seems to be a feasible solution as a substitute. Thus sketch-photo recognition comes into account as a branch of face recognition technique.

*Corresponding author

ICVGIP '10, December 12-15, 2010, Chennai, India

Figure 1: An example of sketch photo pairs taken from AR Database. The original images are cropped into facial part.

Due to great difference between sketches and photos in digital domain, in terms of information content because of their different modalities [see Fig. 1], sketch-photo face recognition is much difficult than conventional photo-photo face recognition. The key for sketch-face photo recognition is to first reduce this modality. During last few years researchers address this problem in two different approaches. One is to generate complex sketch images like pencil sketches from original face images. Henceforth the problem of sketch-photo recognition becomes sketch-sketch recognition. Another approach deals with generating synthesize photos from sketch images. As photo contains more information, it seems to be a good choice for face recognition technique. In that case we can incorporate more information in synthesized photo.

Research on photo synthesis from sketch image is still at its initial stage. Tang et al. [?] are the first to address the problem of face-sketch synthesis and recognition with a significant amount of dataset. They proposed a sketch synthesis method based on eigen transformation algorithm. They used separate transformation for texture and shape to synthesize sketch from face photo. They have taken eigen-transformation with the assumption that the transformation between photo and sketch is linear. They designed a Basian classifier combining texture and shape features to recognize the probing real sketch from the synthesized pseudo-sketches.

Liu et al. [?] used a nonlinear approach for face sketch synthesis. They used a neighborhood-preserving manifold learning method called locally linear embedding(LLE) [?], to synthesize sketch from original face images and they adopted a KNDA based nonlinear discriminative classifier for sketch recognition.

A local strategy based Embedded Hidden Markov model (E-HMM) and selective ensemble (SE) for face sketch synthesis was proposed by Gao et al. [?]. To model the nonlinear relationship between a photo-sketch pair patch they used E-HMM to generate a series of pseudo-photo patches based on several learned models for a given photo patch. Those are integrated using SE strategy to synthesize a finer pseudo-sketch patch and then combined them to get near exact pseudo-sketch patch.

Wang et al. [?] proposed a face photo-sketch synthesis and recognition method using a multi-scale Markov Random Fields (MRF) model. They interpreted recognition technique in both direction : 1) given a face photo, synthesizing a sketch drawing and 2) given a face sketch drawing, synthesizing a photo. They evaluated six different appearance-based face recognition methods and among them Random Sampling LDA (RS-LDA) performed best.

All the previous work for sketch/photo synthesis is done so far either based on global transformation like eigen-face analysis or completely local face sketch synthesis learning like MRF or E-HMM. However, a photo/sketch consists of some global parameter shape as well as some local parameter like texture of a small patch. Therefore combination of both of them can only give good result for all sketch images.

In this paper we build a novel sketch-photo synthesis technique. We use some image warping technique with some control points to get rid of global shape variation from face to face and then built a neighborhood-preserving block matching technique using LLE and a neural network to learn sketch-photo locally. Control points to represent shape of test/query sketch is obtained through well-known ASM. Rest of the paper is organized as follows : section 2 describes the proposed method in detail. The concise algorithm is given in subsection 2.5 and for completeness of description, ASM is briefly described in subsection 2.1. Experimental results and discussion are given in section 3 and 4 respectively. Concluding remarks are presented in section 5.

2. PROPOSED PHOTO SYNTHESIS METHOD

Let I_s be a face sketch image and I_p be the corresponding face photo image, and there are M such pairs in the training dataset. The procedure for generating a pseudo photo from a sketch is equivalent to setting up a relation between a sketch and photo, i.e. $I = P(I_s)$. Intuitively P should be a complex nonlinear mapping. We represent shape of each face object on both sketch and photo images by a set of landmark points (sometimes called as Annotation points or control points). This points are chosen manually in some key positions like at corner and object boundary. Therefore, shape of a face in an image can be represented by some landmark points. In section 2.2 we warp each sketch-photo image pair to mean shape of sketch to get shape free face images. Then in section 2.3 we analyze the local color and texture variation to synthesize shape-free image.

We build a multi-resolution Active Shape Model (ASM) for automatic annotation of landmark pixels for a test sketch image. These annotation points is used for warping the test sketch to mean shape. We also build a neural network and train the sketch-photo control point pairs aligned to the mean shape to get the control points of actual photo.

(a) (b)

Figure 2: An examples of sketch photo pairs from CUHK Database after ploting the annotation points on it.

Figure 2. shows an example of sketch photo pairs from CUHK database after plotting 53 annotation points on it.

2.1 Active Shape Model

Active Shape Model [?, ?] is a statistical model of the shape of objects in training image which iteratively deform to fit to an example of the object in a new image. It captures the natural variability within a class of shapes. First the model is built by learning the patterns of variability from a training database. Here, the training data associated with every training image constitute a sufficiently large set of control points which are chosen based on their location and the application of the model.

2.1.1 *Alignment of Training Set*

Next step is aligning the training set to a common coordinate system in order to be able to compare equivalent points from different shapes. This is achieved by rotation, scaling and translation so that their shapes correspond to a reference shape as closely as possible. A weighted sum of squares of distances between equivalent points on two different shapes is minimized to measure the correspondence as given by

$$E_j = (x_i - T(s_j,\theta_j)[x_j] - t_j)^T W (x_i - T(s_j,\theta_j)[x_j] - t_j) \quad (1)$$

Where x_i and x_j are $2n$ dimension vectors describing n points in i^{th} and j^{th} shapes respectively, $T(s_j,\theta_j)[x_j]$ is a matrix that has the scaling and rotation parameters , t_j is translation matrix and x_i is assumed to be reference frame. W is a diagonal weight matrix for each point. Weights are chosen to give more significance to the points that have lesser variation in a shape.

With this approach, all the shapes are first aligned to the first shape and an initial mean shape is calculated. Now, the shapes are aligned to the normalized mean shape and shapes are re-aligned to the mean shape and a new mean is calculated. The process is iterated until it converges, i.e. sum of all E_j does not change significantly.

2.1.2 Statistical Analysis of the Training set

Once the mean shape is obtained, the deviation $dx_i = x - \hat{x}$ of each shape from the mean shape is calculated from which a covariance matrix is calculated as follows

$$S = (1/M) \sum_{i=1}^{M} (dx_i)(dx_i)^T \quad (2)$$

Where M is the number of shapes.

Now Eigen vectors p^k and Eigen values $\lambda_k (k = 1, 2, ..., 2n)$ of covariance matrix S is calculated such that

$$Sp^k = \lambda_k p^k, \text{ for } k = 1, 2, ..., 2n \quad (3)$$

Then the Eigen values λ_k give the modes of variation of the points x_i. Out of the $2n$ eigen vectors, only t are chosen so that they represent adequate variation described by the t largest eigen values λ_k. By taking the mean shape and adding a linear combination b of the Eigen vectors, it is possible to generate any possible shape x in the domain described by the training data so that

$$x = \hat{x} + Pb \quad (4)$$

where P is the matrix of size $2n \times t$ consisting of t eigen vectors corresponds to t largest eigen values and b is a vector of weights. We assume that each element b_i of vector b individually follows Gaussian with zero mean and standard deviation $\sqrt{\lambda_i}$, then $-3\sqrt{\lambda_i} < b_i < 3\sqrt{\lambda_i}, \forall i = 1, 2, ..., t$.

2.1.3 Search Model

Once the shape variation method from the training set is modeled, next task is to find the instance of the model which suits a given test image in the best possible way. This involves finding the shape and pose parameters which cause the model to coincide with the structures of interest in the image. An instance of the model is given by

$$X = T(s, \theta)[x] + X_c \quad (5)$$

where, $X_c = (x_c, y_c, x_c, y_c, ..., x_c, y_c)^T$ and $T(s, \theta)[]$ is a rotation by θ and scaling by s, and (x_c, y_c) is the position of center of the model frame.

An iterative way is to be used to find this best fit starting from a rough starting approximation. The current instance X of the model is placed on the test image and a region of image around each model point is examined to determine a suggested displacement for that point to give a better fit. One approach to do so is to generate potential images, possibly one for each model point, describing the likelihood of each point in the image being the model point. Adjustments to each point position can then be derived from the gradient of the potential image at the current estimate of the point's position.

The next task is to adjust the pose and shape parameters of the model to move the points from their current locations in image frame X to be as close as possible to the suggested new locations $X + dX$. This is achieved by finding translation dx, rotation $d\theta$ and scaling factor $(1 + s)$ to update the pose variables which best map the current set of points X onto the set of points $X + dX$ by minimizing the same weighted sum of squares as given in equation (1). After this affine transformation, the residual adjustments are done which deform the shape of the model. These are the adjustments dx in the local co-ordinate frame required to cause the points X to move by dX when combined with the effect of the new scale, rotation and translation parameters such that

$$T(s(1+ds), (\theta+d\theta))[x+dx] + (X_c+dXc) = (X+dX) \quad (6)$$

In order to apply the shape constraints, dX is transformed into model parameter space, giving db, the changes in model parameters required to adjust the model points as closely to dx as is allowed. We wish to find db such that

$$x + dx \approx \hat{x} + P(b + db) \quad (7)$$

subtracting (4) from (7) gives

$$dx \approx P(db)$$

so

$$db = P^T dx \quad (8)$$

since $P^T = P^{-1}$, as the columns of P are mutually orthonormal.

Now the calculated changes are used to update the parameters in an iterative scheme at iteration τ :

$$X_{c\tau} \rightarrow X_{c\tau} + w^t dX_{c\tau} \quad (9)$$

$$\theta_\tau \rightarrow \theta_\tau + w^\theta d\theta_\tau \quad (10)$$

$$s_\tau \rightarrow s_\tau + w^s ds_\tau \quad (11)$$

$$b_\tau \rightarrow b_\tau + w^b db_\tau \quad (12)$$

where w^t, w^θ, w^s and w^b are scalar weights. It is ensured that the model deforms into shapes consistent with the class by placing limits on values of b_k . A shape can be considered acceptable only if its Mahalanobis distance D_m is less than a suitable constant D_{max}.

Figure. 3 shows an example of automatic annotation of sketch image from CUHK database using ASM with 53 annotation points. Initially we have given some rough translation with rotation zero and shape coordinate points as mean shape coordinate points as shown in Figure 3.(a). Figure 3.(b) shows the positions of shape coordinate points after 5^{th} iteration and Figure 3.(c) shows its position after convergent.

After convergence, the shape coordinate points are taken as the annotation points of the test sketch. These points are used for warping it to its mean shape and finding out the actual coordinates of those control points in synthesized sketch using a neural network as described in section 3.4.1.

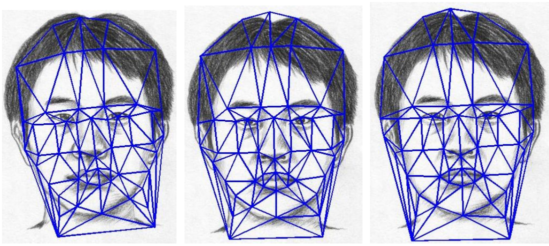
2.2 Generating Shape free face images

To get rid of global variation, we warp all the sketch-photo image pair in the training dataset to its mean shape. For the mean sketch shape obtained from active shape model, represented by some pixel coordinate, we compute a piecewise affine transformation for each image in training dataset by its landmark points and apply this function on it for warping around the image to its mean shape.

2.2.1 Image Warping

Suppose we wish to warp [?, ?] an image I, so that a set of n control points x_i are mapped to new positions, $x_i{}'$. We require a continuous vector valued mapping function $\mathbf{f}$, such that

$$\mathbf{f}(x_i) = x_i{}', \forall i = 1, 2, ..., n \quad (13)$$



(a) (b) (c)

Figure 3: An example of autometic annotation of sketch image using ASM (a) Initial triangulation of annotation points on sketch image (b) After 5 iteration (c) After convergence of ASM.

Given such a function, we can project each pixel of image I into a new image I'. In practice, in order to avoid holes and interpolation problems, it is better to find the reverse map, f', taking ${x_i}'$ into x_i. For each pixel in the target warped image, I' we can determine where it came from in I and fill it in. In general $\mathbf{f}' \approx \mathbf{f}^{-1}$, but is a good enough approximation. Below we consider the forms of $\mathbf{f}$ as piecewise affine interpolator. Note that we can often break down $\mathbf{f}$ into a sum,

$$\mathbf{f}(x) = \sum_{i=1}^{n} f_i(x){x_i}' \tag{14}$$

where each of the n continuous scalar valued functions f_i each satisfy

$$\begin{aligned} f_i(x_j) &= 1 \text{ if } i = j \\ &= 0 \text{ if } i \neq j \end{aligned}$$

this ensure $\mathbf{f}(x_i) = {x_i}'$.

We use simple warping function where it is to assume that each f_i is linear in a local region and zero everywhere else. For instance, in the one dimensional case (in which each x is a point on a line), suppose the control points are arranged in ascending order ($x_i < x_{i+1}$). We would like to arrange that f will map a point x which is halfway between x_i and x_{i+1} to a point halfway between ${x_i}'$ and ${x_{i+1}}'$. This is achieved by setting

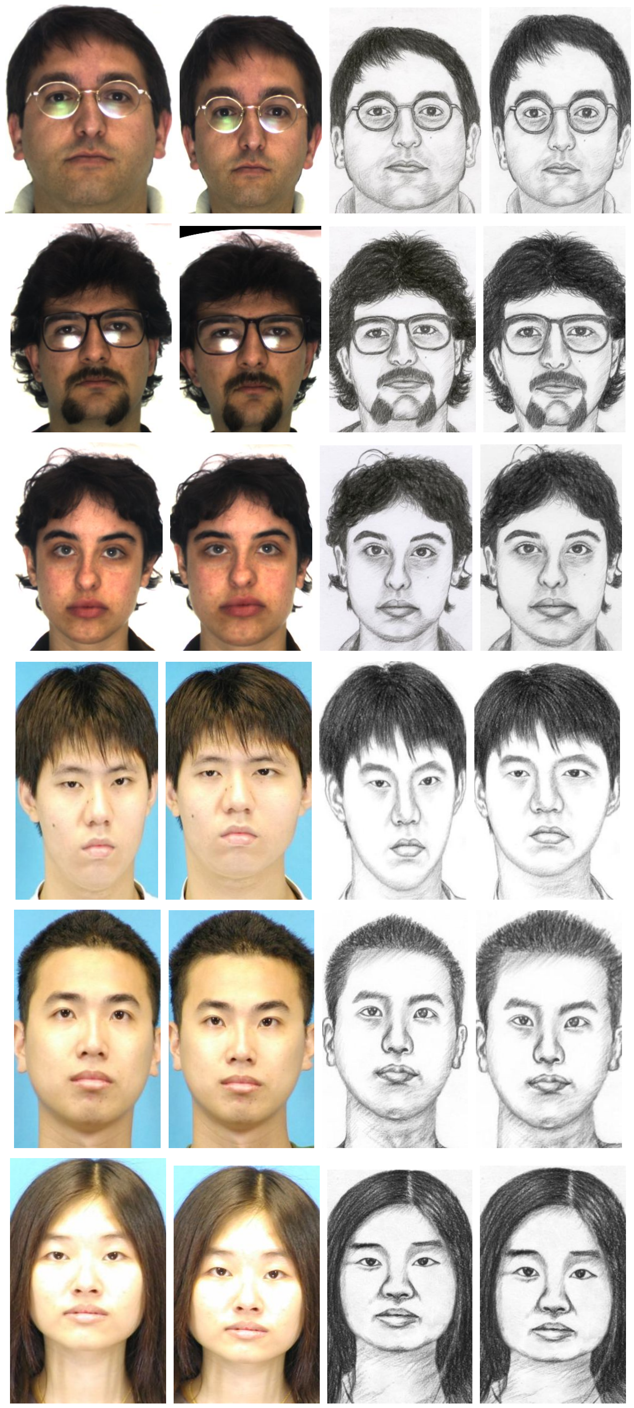
$$\begin{aligned} f_i(x) &= (x - x_i)/(x_{i+1} - x_i) \text{ if } x \in [x_i, x_{i+1}] \text{ and } i \leq n \\ &= 0 \text{ otherwise} \end{aligned}$$

We can only sensibly warp in the region between the control points, $[x_1, x_n]$. In two dimensions, we can use a triangulation (e.g. Delauney) to partition the convex hull of the control points into a set of triangles. To the points within each triangle we can apply the affine transformation which uniquely maps the corners of the triangle to their new positions in I'.

Suppose x_1, x_2 and x_3 are three corners of such a triangle. Any internal point can be written

$$\begin{aligned} x &= x_1 + \beta(x_2 - x_1) + \gamma(x_3 - x_1) \\ &= \alpha x_1 + \beta x_2 + \gamma x_3 \end{aligned}$$

where $\alpha = 1 - (\beta + \gamma)$ and so $\alpha + \beta + \gamma = 1$. For x to be inside the triangle, if $0 \leq \alpha, \beta, \gamma \leq 1$. Under the affine transformation, this point simply maps to



(a) (b) (c) (d)

Figure 4: An example of sketch photo pairs taken from AR Database and CUHK Database, when they are warped around the mean shape.(a) and (c)Original face photo and sketch, (b) and (d) Images after warping around it's mean shape coordinate.

$$x' = \mathbf{f}(x) = \alpha x_1{}' + \beta x_2{}' + \gamma x_3{}' \tag{15}$$

To generate a warped image we take each pixel, x in I, decide which triangle it belongs to, compute the coefficients α, β, γ giving its relative position in the triangle and use them to find the equivalent point in the original image, I. We sample from this point and copy the value into pixel x' in I'.

Note that although this gives a continuous deformation, it is not smooth. Straight lines can be kinked across boundaries between triangles.

Let $2n$ dimension vectors S_s^r and S_p^r represents shapes of a sketch and corresponding photo respectively. Let $\hat{S}_s$ represent mean sketch shape. Let T_s^r and T_p^r be continuous vector valued mappings described as above, which map I_s^r to shape-free sketch $I_s^r{}'$ and I_p^r to shape-free photo $I_p^r{}'$ respectively. i.e.

$$I_s^r{}' = T_s^r(I_s^r)$$

$$I_p^r{}' = T_p^r(I_p^r) \; \forall r = 1, 2, ..., M.$$

satisfying $T_s^r(S_s^r) = \hat{S}_s$ and $T_p^r(S_p^r) = \hat{S}_s$.

Here shape of $I_s^r{}'$ for all the sketches and shape of $I_p^r{}'$ of all the photos have the same shape as mean shape $\hat{S}_s$, only texture are different.

We crop the images both sketch and photo as the bounding box of the mean shape for further processing.

In Figure 4, we shows an example of sketch photo pairs when they are warped around the mean shape. Figure 4(a)-(c) are the original photo-sketch pairs and Figure 4(b)-(d) are the wrapped photo-sketch pairs.

For a test sketch we apply same procedure to get the shape free face sketch. Then we use locally linear embedding (LLE) block matching technique to synthesize the shape free face photo.

2.3 The shape free Pseudo-Photo Synthesis

In this section, we describe our approach to shape-free photo synthesis from the shape-free test sketch based on local patches. We use a local geometry preserving method to learn the mapping relation between shape-free photos and sketches from a training set. LLE is a promising unsupervised manifold learning method, and is widely used for nonlinear dimension reduction of high-dimensional data and image analysis.

The basic idea of LLE [**?**] is to compute neighbor-preserving mapping between a high-dimensional original data space and a low-dimensional feature space, based on simple geometric intuition that each data and its neighbors lie on or close to a locally patch of the manifold. Based on this notion, we present a method, that preserves of local geometry preserving for pseudo-sketch synthesis, with the help of training shape-free photo and sketch image pairs, $I_p^r{}'$ and $I_s^r{}'$, $r = 1, 2, ..., M$.

To reduce the different modalities of sketch-photo and complexity of face structure in training shape-free data, we use popular patch-based strategy [**?**, **?**, **?**, **?**]. We divide the photo and sketch images into N small overlapping and equal size image patches. We denote photo and sketch image patches as g_p^t and g_s^t, $t = 1, 2, ..., M$. Now for a shape-free test sketch image patch g_s, we find out its K nearest searching patch $g_s^{i_k}, i_k = 1, 2, ..., K$ from training samples $T_s^r{}'$ (we will say these block as neighboring blocks). Here we have used euclidean distance as similarity measure. Correlation could also be used as similarity measure. We take corresponding K photo patch $g_p^{i_k}$ and calculate the reconstructing coefficient $w_{i_k i_l}, (i_k, i_l = 1, 2, ..., K)$ for each photo retrieved patch when $i_k th$ patch is fit by the rest of the photo patches $i_l th$, where $i_l = 1, 2, ..., K$ and $i_l \neq i_k$. i.e. each patch $g_p^{i_k}$ compute the reconstruction coefficients $w_{i_k i_l}$ of rest of the photo patches $g_p^{i_l}$ with leave one out method which minimize the error of reconstruction $\epsilon_{i_k}^t(w)$. Where $\epsilon_{i_k}^t(w)$ is defined as

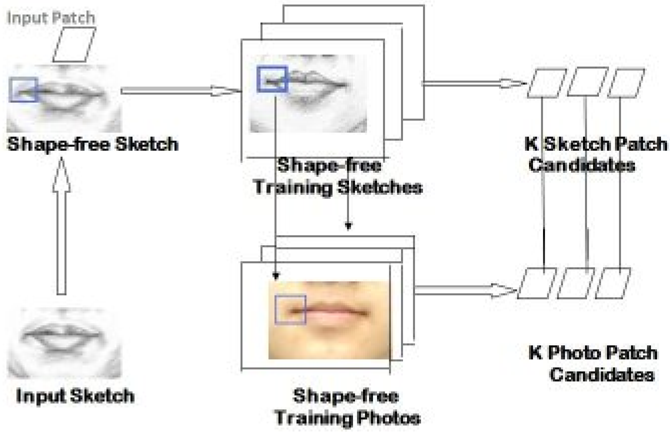


Figure 5: Procedure for finding candidate K neighboring photo patches from training dataset.

$$\epsilon_{i_k}^t(w) = \|g_p^{i_k} - \sum_{l=1}^{K} w_{i_k i_l} g_p^{i_l}\|^2 \; for \; k = 1, 2, ..., K. \tag{16}$$

subject to $\sum_{l=1}^{K} w_{i_k i_l} = 1$, and $w_{i_k i_l} = 0$ if $i_k = i_l$. This is a constrained least square problem. By defining a $K \times K$ matrix Q as

$$Q_k(m, n) = (g_p^{i_k} - g_p^{i_{l_m}})^T (g_p^{i_k} - g_p^{i_{l_n}}) \; for \; k = 1, 2, ..., K. \tag{17}$$

Let $R_k = Q_k^{-1}$, then the above constrained least square problem has the following solution :

$$w_{i_k i_l} = \frac{\sum_{m=1}^{K} R_k(l, m)}{\sum_{i=1}^{K} \sum_{j=1}^{K} R_k(i, j)} \tag{18}$$

where $l = 1, 2, ..., K$, and for each $k = 1, 2, ..., K$

Then we take $\sum_{l=1}^{K} w_{i_k i_l} x_p^{i_l}$ as the estimated synthesized photo patch where k is the index corresponds to minimum error of reconstruction $\min_k(\epsilon_{i_k}^t)$.

The photo synthesis algorithm can be summarized by following major steps :

- step 1. We divide the shape-free test sketch image into some overlapping patches g_{s_j}, $s_j = 1, 2, ..., N$, of equal sizes.
- step 2. For each patch g_s, find its corresponding K nearest sketch patches $g_s^{i_k}$, $k = 1, 2, ..., K$, $i_k \in [1, 2, ..., M]$ from shape-free training sketches $I_s{}'$.
- step 3. For each photo patch $g_p^{i_k}$ corresponding to $g_s^{i_k}$ compute the reconstruction weights $w_{i_k i_l}$ of rest

of the photo patches $g_p^{i_l}$ with the minimum error of reconstruction $\epsilon_{i_k}^t(w)$.

- step 4. Then we assume pseudo synthesize photo patch as $\sum_{l=1}^{K} w_{i_k i_l} x_p^{i_l}$, if k^{th} patch corresponds to minimum reconstruction error, by preserving the local geometry.

- step 5. The above three steps is done for all the sketch patches. In oder to get the local compatibility and smoothness between adjacent synthesized patches we take an average of overlapping region.

We apply the above algorithm for different size image patches and make an average of all of them to get resultant shape free synthesize face photo image.

2.4 Original shape photo synthesis

The synthesized shape-free photo would have the same shape as the mean sketch shape. Therefore, we need to warp it to its correct shape. We apply inverse transformation explained in section 3.2 for image warping, to get back to its original shape. We make a neural network to predict the correct shape of synthesized face photo image from its given sketch.

2.4.1 Neural Network Construction for training Shapes of Photo-Sketch pair

As sketch of a face and its photo should have same shape, they must be functionally related. So we can train shape of each sketch-photo pair from the training database, so that whenever shape of a test sketch is given we can predict shape of its synthesized photo. After aligning every photo sketch pair (S_s^r, S_p^r) of the training database to the mean sketch shape $\hat{S}_s$, we train these pair using feed-forward Multilayer Perceptron Model (MLP) after normalizing them into [0, 1]. We have taken single hidden layer consists of 30 hidden nodes and a sigmoid function as activation function.

We align the shape of the test sketch obtained from ASM to $\hat{S}_s$ and then simulate it to the trained neural network to get the coordinate of shape of original face photo.

We apply the inverse piece-wise affine transformation using the coordinates obtained from the neural network for image warping to get original shaped face synthesized photo.

2.5 Concise Algorithm

In this section we present an overview of the whole algorithm stated above for synthesizing a pseudo face photo from a face sketch.

- Step 1. Model building from the Training dataset :
  (a) Initially build an ASM using the control points of sketch images from training database and estimate the mean shape of sketches [sec. 2.1].
  (b) Warp all the sketch-photo pair images of training dataset to the mean shape (obtained from (a)) described in section 2.2 to generate shape free photo-sketch pair having identical shape.
  (c) Build an MLP to train the control points of photo-sketch pairs, aligned to mean shape of sketches of training database [sec. 2.4.1].

- step 2. Synthesizing pseudo photo for an test sketch :
  (a) Automatic annotation of control points : We apply ASM on test sketch to find out the control points on sketch images.
  (b) Using those control points we warp the test sketch to its mean shape (obtained from step 1.(a)) to get shape free sketch.
  (c) We synthesize shape free pseudo photo image from shape free test sketch using block matching technique as described in section 2.3.
  (d) Align shape coordinate obtained from step 2.(b) to mean shape and feed to trained MLP described in section 2.4.1, to get the estimated control points of output synthesized photo.
  (e) Warp shape free synthesized photo to the actual control points obtained from the neural network to get output synthesized photo.

3. EXPERIMENTAL RESULTS

The experiments are performed on the database containing color face sketch-photo pairs. We have tested our experiment on CUHK Database and AR Database. There are 34 female and 52 male face sketch-photo pairs in the training set and 20 female and 80 male face sketch-photo pairs in the testing set for CUHK database. In AR Database there are 123 face sketch-photo pairs. Among them we have chosen 10 randomly as test set and rest as training set.

The figure. 6 shows the results on AR database. We make 35 annotation/control points manually on each sketch-photo pairs on training dataset. Initially a multi-scale Active Shape Model is constructed on Training sketch images using those annotation points as described in section 3.1. Here we used only three level of training and the initial translation are given using a GUI and initial rotation as 0. We build a MLP with a single hidden layer with sigmoid function as activation function and trained the shape coordinates of each sketch-photo pairs of training dataset align to its mean shape. We tested MLP with different number of hidden nodes in the hidden layer and got best result with 30 hidden nodes. The estimated coordinates of shape obtained from ASM for a given input sketch are used to warp it into its mean shape to get shape free sketch image and feed to the constructed MLP after align into the mean shape to get the actual shape coordinate points of original Photo.

Training face sketch-photo pairs are also warped with respect to mean shape to get the shape free sketch-photo pairs. Then for the shape free input sketch image, we synthesize shape free pseudo image using a manifold learning block matching technique LLE as described in section 3.2. We have taken overlapping blocks sizes as $m \times m$ and overlapping size as one third of the total size. To get the smooth continuous pseudo synthesize photo we apply the model for $m = 12, 18$ and 24, and average them. Then the original shaped pseudo image is generated by warping around the image to the actual shape coordinate obtained from the neural network. Here different size blocks are taken and make average on output images to get rid of blocking effects. Synthesized shape free pseudo photos and actual shaped pseudo images are shown in Figure 6.(c) and Figure 6.(d) respectively.

Figure. 7 shows some example of photo synthesis on CUHK database. Here we have taken 53 annotation points for training sketch-photo image pair. We observe that more annotation points corresponds to more accurate output image. However 30-50 annotation points are good enough to get desirable output. Here we have trained and tested male and female images separately. Output synthesized photo are displayed in Figure 7.(d).

Figure 8. shows an example of photo synthesis where Wang, Tang[**?**] method fails due to variation of shape in sketch image but our procedure perform well for all images.

4. DISCUSSIONS

As it is essentially impossible to mark large number of annotation points manually for test sketches, a multi-scale ASM is constructed for automatic detection of control points and then using neural network we estimate the actual control points of synthesized photo. Since we applied a LLE block matching algorithm on shape free patch, output is really impressive on every image on test dataset, whereas other block matching or eigen transformation technique can give good results on few images. If there is significant amount of variation of shape in face for an input test sketch from average shape, usually other technique fails. But those shape variations are carefully handled here using image warping. Moreover, we are one step ahead towards the sketch-photo recognition technique, as we have applied ASM on sketch images. One can also build another Active Appearance Model (contains shape and texture parameters) on photo database and compute the parameters of the models for each sketch-photo testing pair. Then we can retrieve the original photo from the parameters of ASM for an input sketch and AAM of synthesized photo.

As our proposed method is based on non-iterative technique, it takes very less time compared to E-HMM based or MRF based technique. Moreover, we don't need image registration of test sketch unlike other technique. It is totally automated, except the initial translation of mean shape for ASM and it can be handled using a GUI by moving the cursor around on the face sketch image.

5. CONCLUSIONS

We propose a novel photo generation system from a sketch image. It has potential applications in law enforcement and other photo synthesis and searching application. Based on local geometry preserving manifold learning method LLE and control coordinate points of shape generated by ASM, a synthesized photo of sketch can be automatically generated with the help of training sketch-photo pair samples. Key step of photo synthesis from the test sketch is the novel neighborhood preserving block matching using image warping. Experimental results show the effectiveness of the proposed method.

We could not generate skin color from sketch images, and the database used in this work don't have much variations on skin color, so output synthesized photo would have almost average skin color.

For synthesizing a photo, we need a complete sketch image. So synthesizing a photo from a simple incomplete sketch containing some curves could be an encouraging future work.

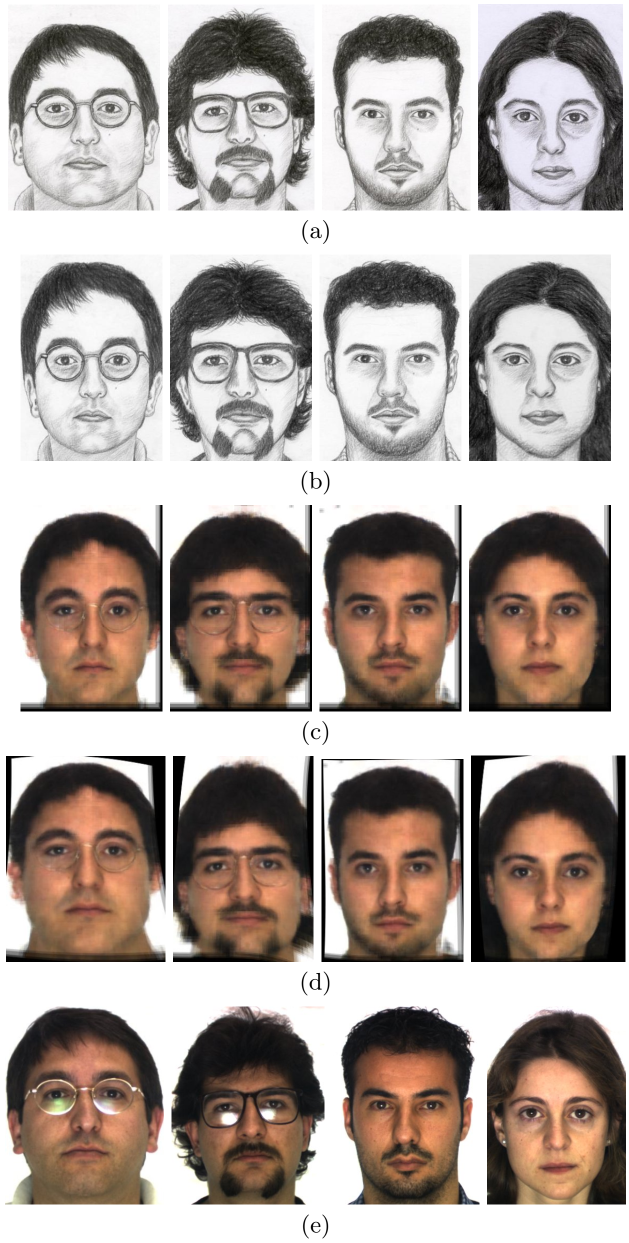


Figure 6: Some photo synthesis results on AR Database (with 35 control points) : (a)Original sketch images drown by artists, (b) After image warping into mean shape (c) shape-free synthesized photo, (d) Output Synthesized photo after image warping into original coordinates and (e) Original Image face photos

6. REFERENCES

[1] T. Cootes, G. Edwards, and C. Taylor. Active appearence models. *IEEE Transactions on Pattern*

(a)

(b)

(c)

(d)

(e)

Figure 7: Some photo synthesis results on CUHK Database (with 53 control points): (a)Original sketch images drown by artists, (b) After image warping into mean shape (c) shape-free synthesized photo, (d) Output Synthesized photo after image warping into original coordinates and (e) Original Image face photos

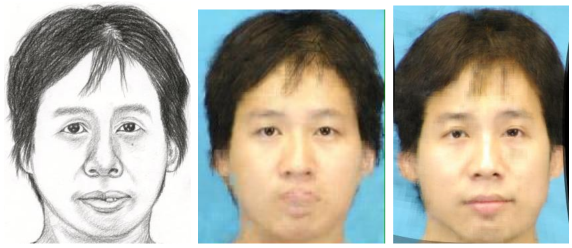


Figure 8: Comparison of synthesized image by proposed method with Wang[?] MRF model procedure.(a) Test sketch (b) Synthesized image by MRF model and (c) Synthesized image by proposed method.

Analysis and Machine Intelligence, 23(6):681–685, June 2001.

[2] T. Cootes and C. Taylor. Statistical models of appearence for computer vision. In *Technical Report*. University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, March 2004.

[3] T. Cootes, C. Taylor, D. H. Copper, and J. Graham. Active shape models - thair training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

[4] X. Gao, J. Zhong, D. Tao, and Xuelong. Local face sketch synthesis learning. *Neurocomputing.*, 71(10-12):1921–1930, February 2008.

[5] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, June 2005.

[6] S. T. Roweis1 and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE.*, 290:2323–2326, December 2000.

[7] D. Ruprecht and H. Muller. Image warping with scattered data interpolation. *IEEE Computer Society Press*, 15(2):37–43, March 1995.

[8] X. Tang and X. Wang. Face sketch synthesis and recognition. *IEEE International Conference on Computer Vision.*, 2(11):1955–1967, October 2003.

[9] J. Wang, H. J. Bao, and W. H. Zhou. Automatic image-based pencil sketch rendering. *Journal of Computer Science and Technology*, 17(5):347–356, May 2002.

[10] X. Wang and X. Tang. Photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 31(11):1955–1967, November 2009.

[11] B. Xiao, X. Gao, D. Tao, Y. Yuan, and J. Li. Photo-sketch synthesis and recognition based on subspace learning. *Neurocomputing.*, 73:840–852, April 2009.