Stochastic Generation of Realistic Content

Umar Mohammed, Simon Prince and Jan Kautz

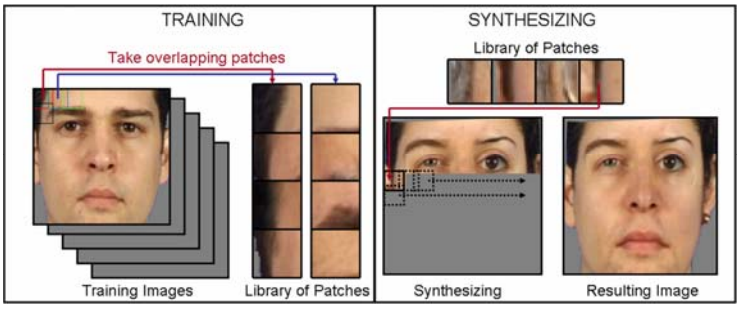
{u.mohammed,s.prince,j.kautz}@cs.ucl.ac.uk Department of Computer Science, University College London

Abstract

Our goal is to generate novel photo-realistic images of a given object class (e.g. faces, trees) using a model trained from example images. To achieve this, we treat training images as samples from a texture with spatially varying statistics and synthesize using a modification of the patch-based method of Efros and Freeman [1]. Unfortunately this generates images that are locally consistent, but globally unrealistic. To resolve this we also learn a weak global model of all the image pixels. This creates images with correct global structure but unrealistic local texture. We demonstrate for the case of faces that combining global and local models allows generation of realistic image content.

Local Model

We can synthesize images which have the correct local face texture using a method similar to [1]. Face images are split into a 9x9 regular grid of overlapping patches. At each image location we build a library of 24000 patches by sampling from 2000 training faces under a variety of small rotations and translations. We synthesize new images piecewise from top-left to bottom-right. At every position, we randomly choose a patch from the appropriate library ensuring visual consistency with existing neighbors. The synthesized images have appropriate local face texture, but are globally unrealistic.



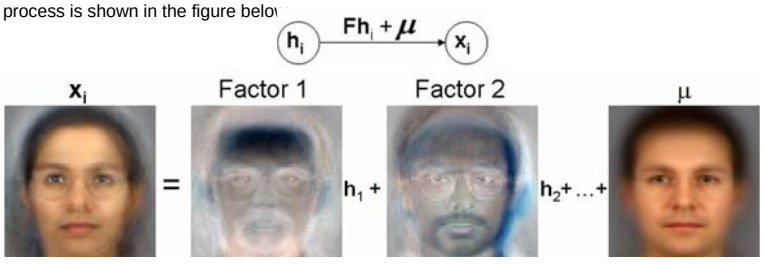
LOCALLY CONSISTENT FACES

Global Model

A global model of faces is learnt using a factor analysis model given by:

$$\mathbf{x}_i = \mathbf{F}\mathbf{h}_i + \mu + \varepsilon_i$$

where $\mathbf{x}_i$ is a face image assumed to have been generated from a point $\mathbf{h}_i$ in a lower dimensional 'face space' by a noisy process. $\mathbf{F}$ is a factor matrix containing the basis vectors of the 'face space', μ is the mean of the training data, and ε is a Gaussian noise term with zero mean and diagonal covariance Σ. The parameters of the model $\theta = [F,\Sigma]$ are learnt using the EM-algorithm. A new face image is then generated by first choosing a random point $\mathbf{h}_i$ and projecting it into image space. The synthesized image is chosen as the peak of the resulting distribution. This process is shown in the figure below:



This is similar to the eigenfaces model, where PCA is used to learn a feature space from a set of face images. New images are then generated by picking a point in feature space and projecting it into image space. The figure below shows example of faces generated using the global model.

GLOBALLY CONSISTENT FACES

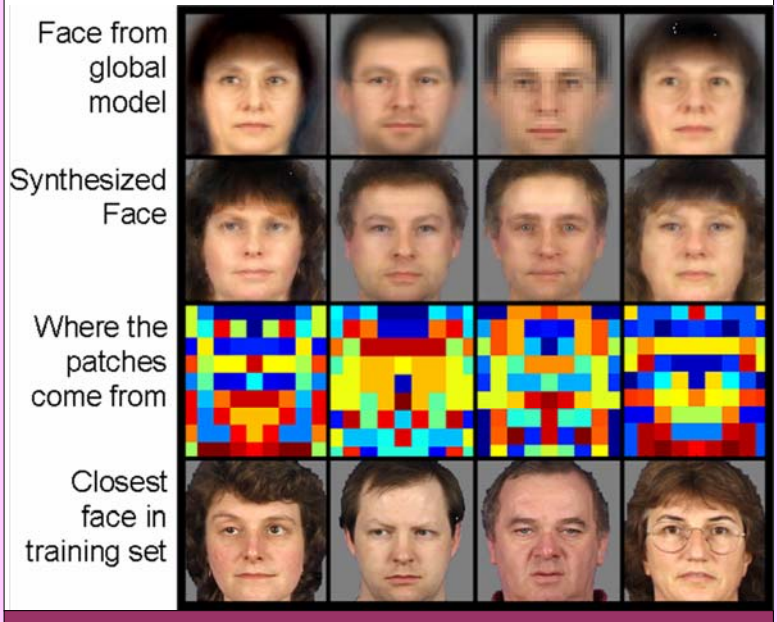
The faces generated using this model are globally coherent but are blurry and do not reproduce realistic local texture.

Combined Model

We generate faces which have a global coherent structure as well as realistic local texture by combining the local and global models. We first generate from the global model, and then sample from the local model where global consistency is ensured by choosing patches that are consistent with each other but also similar to the generated global face. The result of combining the models improves the results from previous methods. However this results in visible seams in patches due to differences in skin tone color.

PLAUSIBLE FACES BUT WITH VISIBLE SEAMS

We remove the seams from the faces using the gradient domain method described in [2]. The final results of the method are shown below.



Changing the style of faces

We can change the style of existing face images. A face is synthesized by replacing the global model with a real face image and sampling from the local model such that all patches are constrained to be of a certain style.

References

[1] A Efros and W Freeman. Image quilting for texture synthesis and transfer. SIGGRAPH 2001, 341-346
[2] P Perez, M Gangnet and A Blake. Poisson Image Editing. SIGGRAPH 2003, 313-318