Domain-specific Progressive Sampling of Face Images

Jianxiong Liu, Christos Bouganis, Peter Y.K. Cheung
Department of Electrical and Electronic Engineering
Imperial College London
London, United Kingdom

***Abstract*—Domain specific knowledge is useful in image processing applications where the target image to process is known to be of a particular image class. It is commonly used as prior knowledge, to model structured image classes such as human faces in order to break limitations posed by various problems. This paper proposes to use domain specific codebook and corresponding sampling patterns learned from example faces, to build a progressive image sampling algorithm specifically for face processing applications. Instead of accessing the whole target face image, the proposed system is able to progressively sample from it and make approximation of it during the process, allowing the process to stop when image quality is considered to have met the requirement. The proposed system is able to identify significant information from the target image and retrieve it at early stage of the sampling, without requiring the target image to be pre-processed as conventional PIT methods do. Therefore it is applicable to situations where such pre-processing is not possible. The experiment shows that the proposed method is able to efficiently sample and reconstruct face images to achieve significant improvement of PSNR over state-of-art method.**

***Index Terms*—Face, progressive sampling, eigenspace**

I. Introduction

Progressive Image Transmission (PIT) is a family of methods that aims to make efficient use of the available bandwidth to transmit large image data [1]. The system can stop at any time during the transmission and still be able to reconstruct an approximation to the ground truth image. Algorithms designed for PIT are able to rearrange the order of transmission so that significant data is transmitted first. The significance is determined by the application and it is most commonly defined as the potential of bringing a high improvement to the quality of reconstructed image.

There are methods in spatial domain and transform domain. The spatial domain methods see the original image stored as pixels been reorganized in such an order that most significant information (such as the most significant bit of each pixel [2]) is transmitted first in the bit stream. Some of the methods [3][4] identify most significant information in the form of pixels and requires the receiver to approximate the ground truth image by interpolation or regression using the pixels received. The transform domain methods on the other hand, transform the original image from spatial domain to frequency domain and transmit significant frequency coefficients accordingly. Discrete Wavelet Transform (DWT) is a popular tool of analysing images and is included by many compression standards, such as the JPEG2000 [5].

There are also works about situations where the statistics of the target image is unknown and pre-processing before transmission required by most PIT methods is not possible. In such situations a model of the image has to be estimated by the receiver and stochastic point sampling methods are often used to blind sample the image. Among various methods in this category, the Adaptive Farthest Point Strategy (FPS) proposed by Eldar et al. [6] is a well defined framework of data-adaptive point sampling method. The method has been studied and improved later by various researchers [7][8]. The estimated model in such methods is refined at each iteration of sampling to approximate the ground truth but is not comparable in performance to PIT methods that can pre-process the target image.

However, when the target image is of a known image class and is therefore well structured, statistics about the image can be derived from examples of the same class. Inspired by the successful use of learned statistics to model human faces, this paper proposes a progressive point sampling method specifically designed for image processing systems that deal with this particular class of images. The core of the propsed method is to use learned codebook and sampled pixels to approximate the target ground truth and break the limitations brought by blind sampling. Unlike conventional point sampling methods, the sampling pattern of the proposed method is learned from a database of example faces. The rest of the paper is arranged as follows: in section II we have an overview of the proposed method; in section III, we explain in detail the learning from a given database of examples; in section IV, experimental results are given showing the capability of the method; section V is the conclusion of the paper.

II. Overview of the Domain Specific Point Sampling of Faces

Unlike stochastic models that are often based on the maximum-distance rule [6] for sampling an unknown image and reconstruct by interpolation or regression, the proposed method benefits from the shared structure of face images, learning both the sampling pattern and the reconstruction codebook from a given collection of examples. The proposed method breaks down face images from the given database into

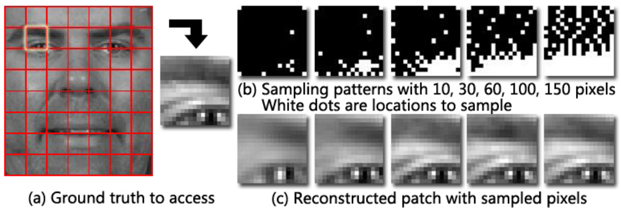

Fig. 1. Patch-based sampling/reconstruction process

small overlapping patches[1] and performs eigen decomposition of the gathered patch examples at each patch location. Sampling patterns of different resolution levels are learned from the statistics of the examples, to suit the reconstruction process. With the learned eigenspace and sampling patterns at each patch location, the system performs point sampling within these patch locations of the target face image and reconstruct in an iterative manner. At every iteration, a patch location with the most validation error is selected to be sampled at a finer resolution, to improve the reconstruction quality. The general process of the proposed method is as follows, and an example is given in Fig.1

Algorithm 1 Overview of the proposed method

Require: for each patch location, the learned eigenspace B and a set of sampling patterns S of different resolution level; target face image I_H to sample to read; an initial sampling of I_H at the lowest resolution level (S_0) of each patch location and the initial reconstruction of each patch

Ensure: the updated reconstruction of the image I_H^*

1: For each patch location, advance to the next sampling resolution level, sample more pixels according to S_1

2: Compute the validation error e of each patch location, defined as the mean squared error of the newly sampled pixels and the previous approximations of these pixels

3: Reconstruct using learned eigenspace at each patch location, using existing samples

4: **while** I_H^* does not meet the user requirement, or the full image has been sampled **do**

5: At the patch location with highest e, advance to the next sampling resolution level by sampling more pixels according to S_{i+1}

6: Update the validation error e by comparing the newly sampled pixels and their approximations before

7: Update the reconstruction of the patch

8: **end while**

9: **return** updated I_H^* as an approximation to I_H

[1]Applying the objective function locally allows the eigenspace projections better flexibility to adapt to local samples, and therefore brings better quality [9] by efficient use of the database with limited amount of example images. More importantly, it makes local updating of the reconstruction during sampling iterations possible, to replace expensive global reconstruction.

III. Design of the Proposed Method

A. Reconstruction by hallucination

The proposed method has the key feature of using learned eigenspace to compensate for the possible reconstruction artefacts caused by point sampling. Therefore here the modeling of the reconstruction is explained first. The reconstruction problem is modeled as a hallucination problem based on the model in [10], with the database being a collection of example faces. In the following we describe the model of the whole image for demonstration purpose. But note that in the proposed method, such model is applied to patches of the image for reasons stated in Sec.II.

Assume the database is organised by Principle Component Analysis into eigenspace B with r dimensions, then given the input ground truth face image I_H the approximation I_H^{opt} is:

$$\mathbf{g}^{opt} = \underset{\mathbf{g}}{arg\min} \|I_H - B * \mathbf{g} - \boldsymbol{\mu}\|_2 \tag{1}$$

$$I_H^{opt} = B * \mathbf{g}^{opt} + \boldsymbol{\mu} \tag{2}$$

Where $\boldsymbol{\mu}$ is the mean face of the database. However, in the context of progressive sampling the input image is not the ground truth but its downsampled version without anti-aliasing filtering. Given a sampled input I_L and sampling matrix S the problem becomes:

$$\mathbf{g}^* = \underset{\mathbf{g}}{arg\min} \|I_L - S * (B * \mathbf{g} + \boldsymbol{\mu})\|_2 \tag{3}$$

Due to the fact that the down-sampling matrix S introduces artefacts and will be constantly refined during the progressive sampling/reconstruction process, we will need extra constraints to enforce prior knowledge. Given a Bayesian treatment the problem of reconstruction is then formulated as *maximum a posteriori* (MAP) problem:

$$\mathbf{g}^* = \underset{\mathbf{g}}{arg\max} [p(I_L \mid \mathbf{g}) * p(\mathbf{g})] \tag{4}$$

With likelihood and prior being:

$$\begin{aligned} p(I_L \mid \mathbf{g}) \quad = \quad & \frac{1}{const} \exp\{-\frac{1}{\sigma_{n,r}^2} [S(B\mathbf{g} + \boldsymbol{\mu}) - I_L]^T \\ & [S(B\mathbf{g} + \boldsymbol{\mu}) - I_L]\} \end{aligned} \tag{5}$$

$$p(\mathbf{g}) = \frac{1}{const} \exp\{-\frac{1}{2} \mathbf{g}^T \Lambda^{-1} \mathbf{g}\} \tag{6}$$

with closed form solution:

$$\mathbf{g}^* = (B^T S^T S B + \sigma_{n,r}^2 \Lambda^{-1})^{-1} B^T S^T (I_L - S\boldsymbol{\mu}) \tag{7}$$

Different from conventional hallucination, the parameter $\sigma_{n,r}^2$ determines the balance between likelihood and prior and is set to take into account different amount of pixels available in I_L: $\sigma_{n,r}^2 = c * \frac{n}{r}$, where c is a constant coefficient and n is the number of pixels currently sampled. Finally, I_H^* can be computed from eq.2 by replacing $\mathbf{g}^{opt}$ with $\mathbf{g}^*$ where the missing pixels can be filled by their corresponding approximations in I_H^*.

B. Learning from database

The ability to hallucinate missing details comes from a well trained base B. The "quality" of the database can be regarded as how much the input image shares similar structure and features as database examples. The quality of database determines the error $\left\| I_H - I_H^{opt} \right\|_2$ of the full projection, and the process of progressively updating $\mathbf{g}^*$ is to approximate $\mathbf{g}^{opt}$ with limited samples. Therefore training from a specific class of images can only be expected to achieve good reconstruction quality of an input image of the same class.

For patch location (i, j) of size $u \times v$, the system preserves the main fraction of the power of eigenvectors trained from N database examples . The number of eigenvectors to preserve $r_{i,j}$ equals to:

$$r_{i,j} = \min r \quad s.t. \quad \sum_{i=1}^{r} \lambda_i > q * \sum_{i=1}^{\min(u \times v, N-1)} \lambda_i \qquad (8)$$

Given the same threshold q, for patches of smaller cross-image variance the number of eigenvectors to preserve will be smaller as well.

C. Sampling order and validation

The sampling process should adapt to the reconstruction process, to increase the information gain per transferred pixels. Data of higher potential priority should be sampled first. To allow for hierarchical refinement over sampling iterations, two types of sampling priorities are defined: patch priority and pixel priority. The task is to identify both the patch and in-patch pixel locations (unsampled sites) that are likely to bring most information gain when sampled in the next iteration.

Priority of patches is computed by the validation error between newly sampled pixels from this iteration and their approximation from previous iteration. Every iteration the system picks the patch location that has the highest average validation error to sample from. The whole process stops when validation errors in all patch locations are below a threshold. Within the patch, we extend the priority score in AFPS [6][7] to answer to a wider variety of reconstruction algorithms:

$$f(x_i, y_i) = d_{i,U} * w_i \qquad U : sampled\ pixel\ locations \quad (9)$$

Where $d_{i,N}$ measures the "distance" from pixel (x_i, y_i) to the current sampled pixels: the likelihood of determining pixel (x_i, y_i) with existing samples. This distance therefore includes but is not limited to Euclidean distance. The weight term w_i is the estimated variance of pixel (x_i, y_i). To accommodate the hallucination process we observe that convergence of the hallucination algorithm is determined by whether or not the pixels of most variance across database examples have been sampled. Therefore we model the two terms as:

$$d_{i,N} = \min_j (1 - corr_{i,j}), \quad (x_j, y_j) \in U \qquad (10)$$

$$w_i = var(I_k(x_i, y_i)), k = 1, 2, ...N \qquad (11)$$

Where $corr_{i,j}$ is the correlation between pixel i and j in database, and I_k is the kth example image in database. An example is given in Fig.2. Notice that when pixel locations with high priority are picked, the priority scores of surrounding locations with high correlation to them are lowered. This reflects the conventional concept of "the distance to the current sampled pixels", but in a form more suitable to the chosen hallucination-based reconstruction algorithm.

The learning process iteratively sets up several levels of sampling patterns according to this priority off-line, with more pixels sampled at higher level (Fig.1). During reconstruction, every time when a patch location is called to be sampled next, the system advances to a higher level of sampling pattern of this patch and samples more pixels accordingly from external data source.

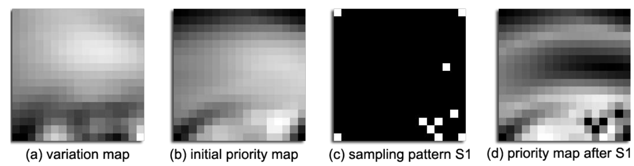

Fig. 2. Learning for sampling patterns: (a) the variation map of the patch location marked in Fig.1; (b) the priority map computed at each pixel by eq.9; (c) sampling pattern (white dots are pixel locations to sample) S_1 at level 1 picks pixel locations with highest priority in (b); (d) updated priority map after S_1.

IV. Experiment Result

The proposed method is evaluated on the FERET database [11] of 1752 frontal faces, and the ORL database [12] of 400 faces. All faces were resized to 129×113 and were broken down to 17×17 patches overlapping each other by 1 column/row. For each patch location, sampling patterns containing 10, 30, 60, 100 and 150 pixels are pre-learned. For each test, a certain number of training face are randomly selected from the database, excluding examples of the test subject. The reported performance below is an average across repeated tests. For demonstration purpose, the sampling process in all tests stops when about 20% of the total pixels are sampled, showing the most informative data during the sampling process. The proposed method is compared with the grid AFPS [7] to show the impact of applying domain specific knowledge in sampling without pre-processing or compressing the target image.

An example of reconstruction is shown in Fig.3. In this particular example, the eigenspace codebook was trained from 500 random faces from the FERET database, excluding face examples of the testing subject. It can be seen that the proposed method can achieve a better approximation quality (in PSNR) than state-of-art method does, especially at early stages. The faces reconstructed by the proposed method also exhibit much sharper features, by virtue of the hallucination based reconstruction algorithm. Even though the training database is randomly selected and does not include examples of the same testing subject, the codebook learned can still resemble the target face by filling in the missing pixels with hallucinated data.

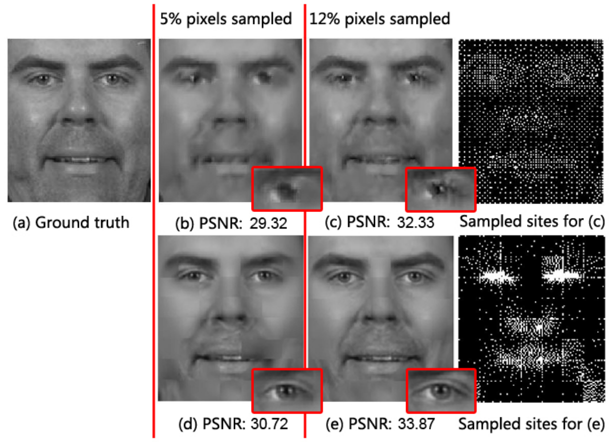

Fig. 3. Example reconstructions with different amount of pixels sampled; (b)(c) are reconstruction examples obtained from global AFPS and triangle-based linear interpolation; (d)(e) are reconstruction examples obtained from the proposed method with $q = 99.9\%$. The locations of sampled sites for (c) and (e) are shown to their right side.

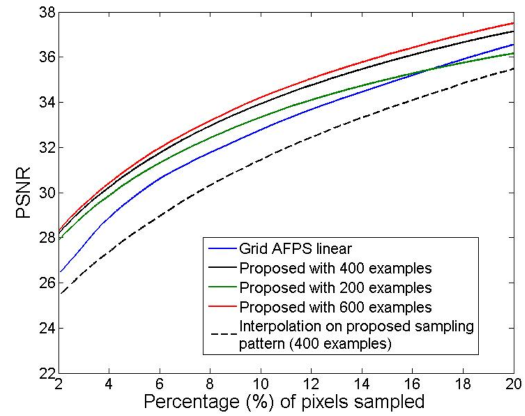

Fig. 4. Performance evaluation with $q = 99.9\%$; for the pacth location in Fig.1, 100, 123 and 134 eigenvectors are preserved for 200, 400 and 600 training examples in the database, respectively.

Fig.4 shows the overall sampling performance of both the reference method and the proposed method. The performance is taken at each sampling iteration given $q = 99.9\%$, and is measured in reconstruction PSNR vs. percentage of samples required. It can be seen that a larger training database allows for a higher flexibility of the projection to fit samples from target image, due to increased number of eigenvectors in the codebook. This difference becomes more significant in late stages of the sampling process, when a large amount of pixels are sampled. Since the hallucination based method is mainly designed to improve the reconstruction quality early on when samples are relatively sparse, a significant improvement in PSNR can be seen in this graph compared with that of the conventional point sampling/interpolation scheme.

Additionally, the dashed line in Fig.4 shows the reconstruction of the target image using triangulation based linear interpolation, but with samples retrieved by the sampling pattern generated for the proposed reconstruction algorithm. It can be seen that different sampling patterns serve different reconstruction algorithms: while the learned sampling patterns improve the convergence rate of the hallucination based algorithm, it is not derived from the continuity assumption of images, which is the foundation of interpolation algorithms. Therefore the good sampling performance of the proposed system comes from both the use of domain specific codebook for reconstruction, and the specially tailored sampling patterns.

V. Conclusion

In this paper, we propose to apply domain specific knowledge in point sampling strategies. A progressive sampling/reconstruction method is introduced to replace conventional point sampling methods in sampling for face image processing system. Same as convential point sampling strategies, the proposed method does not require pre-processing or compressing the target image before transmissin. Instead of using a stochastic model, the proposed method utilize eigenspace codebook and sampling patterns learned from statistics of face examples in a database. The proposed method achieved a improvment both in PSNR of the reconstructed image, and in the visual quality by virtue of hallucination based algorithm.

References

[1] K.-H. Tzou, "Progressive image transmission: a review and comparison of techniques," *Optical Engineering*, vol. 26, no. 7, pp. 267 581–267 581, 1986.

[2] C.-C. Chang, F.-C. Shiue, and T.-S. Chen, "A new scheme of progressive image transmission based on bit-plane method," in *Communications, 1999. APCC/OECC'99. Fifth Asia-Pacific Conference on... and Fourth Optoelectronics and Communications Conference*, vol. 2. IEEE, 1999, pp. 892–895.

[3] S. Rajesh, K. Sandeep, and R. Mittal, "A fast progressive image sampling using lifting scheme and non-uniform b-splines," in *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1645–1650.

[4] L. Demaret, N. Dyn, and A. Iske, "Image compression by linear splines over adaptive triangulations," *Signal Processing*, vol. 86, no. 7, pp. 1604–1616, 2006.

[5] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 36–58, 2001.

[6] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Zeevi, "The farthest point strategy for progressive image sampling," *Image Processing, IEEE Transactions on*, vol. 6, no. 9, pp. 1305–1315, 1997.

[7] Z. Devir and M. Lindenbaum, "Adaptive range sampling using a stochastic model," *Journal of computing and information science in engineering*, vol. 7, no. 1, pp. 20–25, 2007.

[8] C. Moenning and N. A. Dodgson, "Fast marching farthest point sampling for implicit surfaces and point clouds," *Computer Laboratory Technical Report*, vol. 565, 2003.

[9] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *Signal Processing Letters, IEEE*, vol. 18, no. 6, pp. 367–370, 2011.

[10] C. Liu, H. Shum, and W. Freeman, "Face hallucination: Theory and practice," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[11] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *Image and vision computing*, vol. 16, no. 5, pp. 295–306, 1998.

[12] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*. IEEE, 1994, pp. 138–142.