Epitomic image factorization via neighbor-embedding

Mehmet Turkan, Martin Alain, Dominique Thoreau, Philippe Guillotel, Christine Guillemot

EPITOMIC IMAGE FACTORIZATION VIA NEIGHBOR-EMBEDDING

Mehmet Türkan[1,2], *Martin Alain*[1,3], *Dominique Thoreau*[1], *Philippe Guillotel*[1], *Christine Guillemot*[3]

[1] Technicolor R&D, Cesson Sévigné, France
[2] Izmir University of Economics, Izmir, Turkey
[3] INRIA, Rennes, France

ABSTRACT

We describe a novel epitomic image representation scheme that factors a given image content into a condensed epitome and a low-resolution image to reduce the memory space for images. Given an input image, we construct a condensed epitome such that all image patches can successfully be reconstructed from the factored representation by means of an optimized neighbor-embedding strategy. Under this new scope of epitomic image representations aligned with the manifold sampling assumption, we end up a more generic epitome learning scheme with increased optimality, compactness, and reconstruction stability. We present the performance of the proposed method for image and video up-scaling (super-resolution) while extensions to other image and video processing are straightforward.

Index Terms— Epitome learning, image factorization, neighbor-embedding, image up-scaling, super-resolution, compression

1. INTRODUCTION

An *epitome* is considered as a miniature, condensed representation of essential statistical properties of ordered datasets such as matrices representing images and videos, audio signals, text, or genome sequences in molecular biology. The epitome representation is generally much smaller than a summary (e.g., [1]), and it contains many small overlapping parts of the signal with less repetitions –and with some level of generalization– when compared to the compaction of ordinary textures (e.g., [2]). Being such a condensed and generalized version representing high-order statistics contained in the data, epitomes can be used for data mining and other machine learning and signal processing tasks such as texture classification, texture transfer, image inpainting, denoising, super-resolution, compression, segmentation, recognition, and indexing (e.g., [3, 4, 5, 6]).

The concept of epitomes is first introduced in [3] within a patch-based probabilistic framework by iteratively optimizing the factorization of a given image $\mathcal{I}$ into an epitome $\mathcal{E}$ and a smooth mapping $\mathcal{M}$, such that $(\mathcal{E}, \mathcal{M})$ gives rise to an efficient and powerful generative model. This image epitome model is later extended to video epitomes in [4] by means of 3D space-time patch volumes drawn from the input video. In [5], the epitome is formed by learning arbitrary shaped jigsaw structures occurring in the given set of training images. This epitome scheme leads also to a probabilistic generative model with $(\mathcal{E}, \mathcal{L})$ where $\mathcal{L}$ is an associated *offset map* of the same size with, and characterized by, the input image $\mathcal{I}$.

Another family of epitome construction algorithms exploits textural *(self-)similarities* [7, 8] within and across images, e.g., [6, 9]. In these methods, the input image $\mathcal{I}$ is described with a generative model $(\mathcal{E}, \mathcal{T})$ which consists of an optimized epitome $\mathcal{E}$ and a compact transform (correspondence or assignation) map $\mathcal{T}$. The epitome $\mathcal{E}$ is composed of disjoint regions called *epitome charts*, and each epitome chart corresponds to a specific region in the input image. Epitome charts are initialized sequentially, and each of them is extended progressively by retaining only useful image information while optimizing the trade-off between image reconstruction quality and epitome size. In other words, the epitome is constructed by optimally eliminating repeated texture patterns in a given image or set of images. Hence, the first step is to search for (self-)similarities within or across images. One can consider the KLT (Kanade-Lucas-Tomasi) feature tracking algorithm to determine the repeated texture patterns as in [6] where the information about translations, rotation angles, scaling factors, and reflections need to be kept in the transform map $\mathcal{T}$. Alternatively, repeated patterns can be identified in and removed from the image using a simple block-matching (BM) algorithm. In this way, the transform map may contain only translations of full- or sub-pel precision [9]. It has later been shown in [10] that such an optimized epitome itself can be placed in the heart of an image compression scheme leading to very high bit-rate savings.

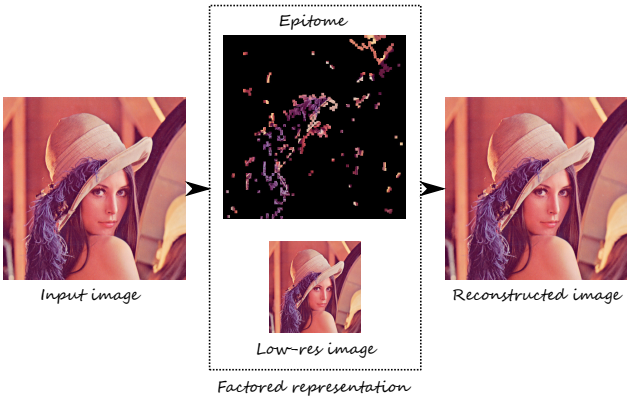

Fig. 1: Factoring the given image content into a condensed epitome and a low-resolution image through neighbor-embedding.

Motivated with all these greedy and promising progress of the epitomes concept and applicability to wide range of signal processing tasks, in this paper we broaden the scope of epitomic image representations. Assuming that most natural images are sampled from low-dimensional (sub-)manifolds, the densely sampled small texture patches can successfully be reconstructed as a weighted linear combination of their nearest-neighboring (NN) patches, i.e., *neighbor-embedding*. This is the same underlying main idea of many manifold learning algorithms, e.g., [11, 12, 13], and with this paper we exploit the manifold sampling assumption by constructing optimized epitomes especially for reducing the memory space for images and video sequences while keeping the highest possible reconstruction quality. As illustrated briefly in Fig. 1, the given input image $\mathcal{I}$ is factored

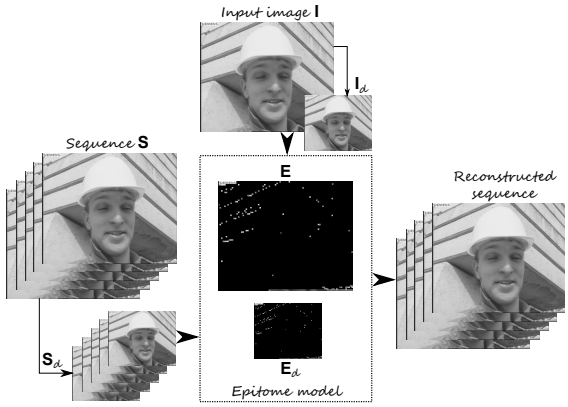

Fig. 2: The epitome model $(\mathbf{E}, \mathbf{E}_d)$ is obtained from images $\{\mathbf{I}, \mathbf{I}_d\}$. This model can be used for other images containing similar content to $\mathbf{I}$, e.g., a set of video frames $\mathbf{S}$ in the same shot.

into a condensed epitome $\mathcal{E}$ and a low-resolution image $\mathcal{I}_d$ resulting in a powerful generative model with $(\mathcal{E}, \mathcal{I}_d)$ for $\mathcal{I}$. An equivalent model $(\mathcal{E}, \mathcal{E}_d)$ can then be used for other images containing similar content to $\mathcal{I}$, e.g., a set of video frames in the same shot (see Fig. 2). Here $\mathcal{E}_d$ denotes a low-resolution epitome obtained by masking $\mathcal{I}_d$.

In short, the epitomic representation described in this manuscript optimizes the up-scaling quality of the image $\mathcal{I}_d$ by means of the model $(\mathcal{E}, \mathcal{E}_d)$ through neighbor-embedding. Generally speaking, neighbor-embedding based up-scaling (super-resolution) algorithms consider local geometric similarities of low-resolution (LR) and high-resolution (HR) patch (or feature) spaces, e.g., [14, 15]. These methods first characterize (literally, learn) intrinsic properties of the LR neighborhood using a set of NN patches, and then transfer this local geometry to the HR patch space. However, the crucial problem arises from the one-to-multiple mapping from LR to HR patch spaces which may completely destroy the initial local geometric similarity assumption for these two corresponding spaces. This is due to the fact that many HR patches can give rise to the same LR patch under the same degradation procedure. Therefore, one needs rather a more compact and generic representation characterizing locally-linear and smooth manifold structures of these patch spaces while keeping their local geometric similarities intact. In the rest of this paper, we describe an efficient solution to this problem resulting in a more generic epitome learning scheme with increased optimality, compactness, and reconstruction stability.

This paper is organized as follows. We first detail the main ideas and steps of the proposed algorithm in Sec. 2. We then present obtained experimental results for image up-scaling in Sec. 3. In Sec. 4, we finally give a conclusion by describing possible extensions.

2. EPITOMES THROUGH NEIGHBOR-EMBEDDING

It is worth starting with strong perspectives of the proposed epitomic representation: (1) relying on the manifold sampling assumption, the constructed epitome model will be able to successfully reconstruct most texture patches extracted from other images with similar content; and (2) relying on a multi-patch scheme rather than a single-patch approach and optimizing the epitome for the whole image by discarding local epitome charts, the constructed epitome is expected to be optimally compact and also be a powerful tool for more stable image reconstruction. In the following, we give the underlying ideas and main steps of the proposed image factorization algorithm.

2.1. Motivation & Notation

Our basic aim is to obtain a generative epitome model $(\mathbf{E}, \mathbf{I}_d)$ which is optimized for the reconstruction of a given input image $\mathbf{I}$. Here $\mathbf{E}$ denotes a condensed epitome which is constructed by retaining only a small subset of useful texture patches in the input image $\mathbf{I}$, and $\mathbf{I}_d$ is the low-resolution image, such that $\mathbf{I}_d = (\mathbf{I} * \mathbf{H})\downarrow_s$, where $\mathbf{H}$ is a low-pass (anti-aliasing) filter, $*$ and $\downarrow_s$ represent 2D convolution and down-scaling (by a factor s in both directions) operators, respectively. Note that s can take any integer or non-integer value provided that one has to carefully relate (irregular) grid points of images $\mathbf{I}$ and $\mathbf{I}_d$ in order to ensure that spatially co-located HR/LR patch pairs extracted from these images are in phase [16, 17].

Above problem is by now converted into an up-scaling optimization on the image $\mathbf{I}_d$ for the reconstruction of $\mathbf{I}$. As stated earlier, we will follow the manifold sampling assumption and extend it for constructing an up-scaling epitome $\mathbf{E}$. For the solution, we choose to apply a variant of neighbor-embedding tool because (a) it leads to a multi-patch scheme where whole patch space can be spanned with a small number of selected texture patches via weighted linear combinations, and (b) it can be generalized to other images which contain similar content, i.e., to input patches with small distortions and translations which fall into the same sub-manifold. Note that for a given model $(\mathbf{E}, \mathbf{I}_d)$, one can obtain an equivalent and *generalized* model $(\mathbf{E}, \mathbf{E}_d)$ where $\mathbf{E}_d$ is a low-resolution epitome in which texture patches (spatially co-located to those of in $\mathbf{E}$) are kept from $\mathbf{I}_d$. Our main idea is to make sure that retained texture patch pairs in $(\mathbf{E}, \mathbf{E}_d)$ strictly obey the local geometric similarity assumption for LR and HR patch spaces so that one can safely transfer intrinsic properties of the LR neighborhood to the HR patch space. As illustrated in Fig. 2, the model $(\mathbf{E}, \mathbf{E}_d)$ can then be used for up-scaling other images containing similar content to $\mathbf{I}$, and it may finally be packed into an *epitome atlas* [6] for a more compact representation.

Let us now specify our notation in this paper. We assume a lexical ordering of image pixels and square image blocks (or patches) as stacked column vectors. We take pixel grid points of the image $\mathbf{I}_d$ as the reference grid and restrict ourselves to integer s values for simplicity. We consider a **block** when the image patches extracted are aligned with the disjoint block-grid structure, and a **patch** when the extracted image patches are aligned with the overlapping pixel-grid structure. As demonstrated in Fig. 3, each LR block $\mathbf{x}'_j, j = 1 \dots J$, of size $m \times m$ extracted from $\mathbf{I}_d$ has always one spatially co-located HR block $\mathbf{y}'_j$ of size $n \times n$ extracted from the block-grid of $\mathbf{I}$ where $n = sm$. For each LR patch $\mathbf{x}_i, i = 1 \dots I$, however, one needs to search for the spatially co-located HR patch $\mathbf{y}_i$ in regular offsets (s-pixels in both directions) from $\mathbf{I}$. Optionally, one might initially interpolate $\mathbf{I}_d$ by a factor of s to be of same spatial resolution with $\mathbf{I}$.

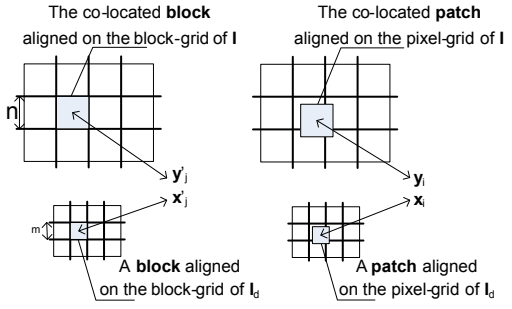

Fig. 3: A block is an image patch aligned with the disjoint block-grid structure, and a patch is aligned to the overlapping pixel-grid.

In this case, HR/LR block and patch pairs will be of same spatial size and pixel-grid of $\mathbf{I}$ can completely be covered. In addition, extracted block-grid patches can also overlap where the overlapping areas may uniformly, or weighted, averaged after reconstruction. Without loss of generality, let us denote the obtained HR/LR block and patch pairs in sets as $\mathcal{B} = \{\mathbf{y}'_j, \mathbf{x}'_j\}_{\forall j}$ and $\mathcal{P} = \{\mathbf{y}_i, \mathbf{x}_i\}_{\forall i}$, respectively. It is important to note here that $\mathcal{B}$ is a subset of $\mathcal{P}$.

2.2. Epitome construction

Given sets $\mathcal{B}$ and $\mathcal{P}$, we start with a zero initialized epitomic model $(\mathbf{E}, \mathbf{E}_d)$ where $\mathbf{E}^0 = \mathbf{0}$ and $\mathbf{E}_d^0 = \mathbf{0}$ are of same spatial size with $\mathbf{I}$ and $\mathbf{I}_d$, respectively. We iteratively select patch pairs from $\mathcal{P}$, and progressively place them (i.e., square patches) in the epitome model according to their index i. In the case where selected patches overlap, one needs to carry out an averaging on the overlapping pixels. The final epitomic model obtained will contain an optimum subset of $\mathcal{P}$ which can reconstruct all block pairs in $\mathcal{B}$ for a given criterion through an optimized neighbor-embedding. In the following, we will omit the notion of *epitome charts* and optimize our epitome for the whole image in order to be optimally compact.

Initialization: Although there might be different ways of initialization, we follow the simplest and most general framework which is also highly compatible with our epitome extension proposal. We select an initial patch pair $\{\mathbf{x}_i, \mathbf{y}_i\}$ from $\mathcal{P}$ such that

$$i = \arg\max_i \sum\nolimits_j \delta\Big(\big\|\mathbf{y}'_j - \mathbf{y}_i\big\|_2^2 \leq \epsilon\Big) \tag{1}$$

where $\delta\,()$ is a binary function which returns 1 when its argument is `true` and 0 otherwise, and ϵ denotes an admissible approximation error. Here basically we initialize our epitome model with a patch pair which reconstructs maximum number of blocks of $\mathbf{I}$. Note that this is a special case which turns out to be a simple BM. As a result at iteration $t = 1$, we obtain $\mathbf{E}^1 = \{\mathbf{E}^0 \cup \mathbf{y}_i\}$ and $\mathbf{E}_d^1 = \{\mathbf{E}_d^0 \cup \mathbf{x}_i\}$.

Extension: To extend the model for $t > 1$, we similarly maximize the number of successfully reconstructed blocks in $\mathbf{I}$ by systematically checking all possible extensions $\{\mathbf{y}_i, \mathbf{x}_i\} \notin (\mathbf{E}^{t-1}, \mathbf{E}_d^{t-1})$ from $\mathcal{P}$ to currently available epitome model such that

$$\left(\mathbf{E}^t, \mathbf{E}_d^t\right)_i = \left(\{\mathbf{E}^{t-1} \cup \mathbf{y}_i\}, \{\mathbf{E}_d^{t-1} \cup \mathbf{x}_i\}\right)$$

where

$$i = \arg\max_i \sum\nolimits_j \delta\bigg(\Big\|\mathbf{y}'_j - \sum\nolimits_k w_k \mathbf{Z}_k^i\Big\|_2^2 \leq \epsilon\bigg), \tag{2}$$

$$\{w_k\} = \underset{\{w_k\}, \sum w_k = 1}{\arg\min} \Big\|\mathbf{x}'_j - \sum\nolimits_k w_k \mathbf{z}_k^i\Big\|_2^2 \;\text{ s.t. }\; \min|k|\,. \tag{3}$$

We now turn our attention to neighbor-embedding, and suppose that the set $\mathcal{Z} = \{\mathbf{Z}_\tau^i, \mathbf{z}_\tau^i\}$, $\tau = 1 \ldots t$, represents available HR/LR patch pairs in the epitome model $\left(\mathbf{E}^t, \mathbf{E}_d^t\right)_i$. The optimum neighbor-embedding weights $\{w_k\}$ in (3) mainly characterize the intrinsic geometry of the local LR neighborhood of $\mathbf{x}'_j$ using its k-NN patches $\{\mathbf{z}_k^i\}$ taken from the set $\{\mathbf{z}_\tau^i\}$, $k \in [1, \min\{t, K\}]$. The *sum-to-one* constraint on these coefficients not only provides invariance to translations but also enforces the reconstruction to lie in the sub-space spanned by these k-NN patches. We further consider a *sparsity* constraint which basically imposes to use as few number of patches as possible for the reconstruction while rejecting the noise, and above all, for obtaining a sparse –hence *compact*– epitome model.

While ignoring the sparsity constraint, one can solve (3) with the help of a Lagrangian form for $\mathbf{w} = [w_1 \ldots w_k]^T$ by [12] as

$$\mathbf{w} = \left(\mathbf{G}^{-1}\mathbf{1}\right) \Big/ \left(\mathbf{1}^T\mathbf{G}^{-1}\mathbf{1}\right) \tag{4}$$

Input: $\mathcal{Z} = \{\mathbf{Z}_\tau^i, \mathbf{z}_\tau^i\}_{\tau=1}^t$, $\mathbf{x}'_j$, K
Output: Selected subset $\{\mathbf{Z}_k^i, \mathbf{z}_k^i\}$ from $\mathcal{Z}$
$k = 0$, $\hat{\mathbf{z}} = [\,]$, $e_0 = \big\|\mathbf{x}'_j\big\|_2^2$
while $k < \min\{t, K\}$ **do**
  $k = k + 1$
  $\tau_k = \arg\min_{\tau \notin \{\tau_1, .., \tau_{k-1}\}} \big\|\mathbf{x}'_j - \left[\hat{\mathbf{z}}\;\; \mathbf{z}_\tau^i\right]\mathbf{w}\big\|_2^2$
  $e_k = \big\|\mathbf{x}'_j - \left[\hat{\mathbf{z}}\;\; \mathbf{z}_{\tau_k}^i\right]\mathbf{w}\big\|_2^2$
  if $e_k \geq e_{k-1}$: $k = k - 1$, **break** while–loop **endif**
  $\mathbf{z}_k^i = \mathbf{z}_{\tau_k}^i$, $\mathbf{Z}_k^i = \mathbf{Z}_{\tau_k}^i$, $\hat{\mathbf{z}} = \left[\hat{\mathbf{z}}\;\; \mathbf{z}_k^i\right]$
end

Algorithm 1: Iterative patch selection. $\mathbf{w}$ is calculated with Eqn. 4 and $\mathbf{1}^T\mathbf{w} = 1$.

where $\mathbf{G}$ denotes the Gram matrix of patches $\{\mathbf{z}_k^i\}$ centered around $\mathbf{x}'_j$ and $\mathbf{1}$ is the column vector of ones. Considering now the sparsity constraint, we need to search for a solution to (3) where the number k of used NN patches is minimized. For this aim, we adopt an iterative patch selection method similar to [18]. At each iteration k, we greedily select one patch $\mathbf{z}_k^i$ (and the corresponding $\mathbf{Z}_k^i$) from $\mathcal{Z}$ and use it with the previously obtained $(k-1)$ patches to minimize the error $\big\|\mathbf{x}'_j - \hat{\mathbf{z}}\mathbf{w}\big\|$ where $\hat{\mathbf{z}} = \left[\mathbf{z}_1^i \ldots \mathbf{z}_k^i\right]$ and $\mathbf{w}$ is calculated with (4). We keep iterating until the reconstruction error energy stops decreasing or $k = \min\{t, K\}$[1]. This method is summarized in Algorithm 1.

We finally check if the intrinsic geometry of the local neighborhood of $\mathbf{x}'_j$ which is characterized by $\{w_k\}$ using $\{\mathbf{z}_k^i\}$ in (3) is similar to the corresponding neighborhood of $\mathbf{y}'_j$ using $\{\mathbf{Z}_k^i\}$ via (2). The main idea is to extend the current epitome with $\{\mathbf{y}_i, \mathbf{x}_i\} \in \mathcal{P}$ which optimally preserves the neighbor-embedding relationship between patch pairs retained in the model while maximizing the number of reconstructed blocks in $\mathcal{B}$. We may repeat this procedure until all blocks are successfully reconstructed with the given criterion. Note that $\mathcal{B} \subset \mathcal{P}$ and thanks to sparsity constraint, the blocks in $\mathbf{I}$ which cannot be successfully reconstructed via neighbor-embedding will be automatically included in the epitome model without destroying the neighbor-embedding relation and leading to a solution with $|k| = 1$. One might also keep track of the whole image reconstruction quality, e.g., peak signal-to-noise ratio (PSNR), or the epitome size as a stopping criterion. The recursively selected and retained patches in the epitome model guarantee a gradual improvement on the reconstruction quality while epitome size increases.

2.3. Image reconstruction

Suppose that we are given an image $\mathbf{I}_d$ and the model $(\mathbf{E}, \mathbf{E}_d)$. We first extract $m \times m$ patches $\mathbf{x}_q, q = 1 \ldots Q$, from $\mathbf{I}_d$, and then upscale them using the available patch pairs $\{\mathbf{Z}_\tau, \mathbf{z}_\tau\}$, $\tau = 1 \ldots T$, in the epitome model. For each given $\mathbf{x}_q$ hence, we obtain a k-NN set $\{\mathbf{Z}_k^q, \mathbf{z}_k^q\}$ using Algorithm 1 and calculate the local neighborhood geometry $\{w_k\}$ using $\{\mathbf{z}_k^q\}$. We finally obtain the corresponding up-scaled patch $\mathbf{y}_q$ of size $n \times n$ by $\mathbf{y}_q = \sum_k w_k \mathbf{Z}_k^q$.

We will implicitly enforce regular overlaps between adjacent patches to avoid blocking artifacts for the final reconstruction. We uniformly average overlapping patches in order to force all image patches to agree on the overlapped areas, hence to satisfy the local compatibility and smoothness constraints. Here one might also envisage other weighting strategies than uniform averaging.

[1] If $t \leq K$, we consider all t patches as NN to $\mathbf{x}'_j$. If $t \gg K$, we only consider κ-NN patches to $\mathbf{x}'_j$ from $\mathcal{Z}$ and input to Algorithm 1, $\kappa \geq K$.

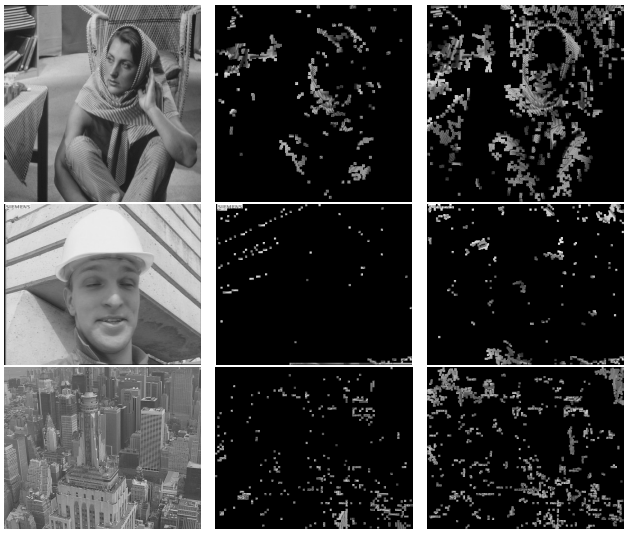

Fig. 4: (Left-to-right–each row) Test image, epitome obtained with our method and the method in [9] where both epitomes (having different models) reconstruct the input test image at the same quality.

3. EXPERIMENTAL RESULTS

We report our experimental results for the *Barbara* (512) image, and *Foreman* (CIF) and *City* (4CIF) sequences. For simplicity $s = 2$, and we apply and test the proposed algorithm on the luma (Y) component only. We calculate epitome models for *Barbara* and the first frames of *Foreman* and *City* sequences. For video sequences, we use the corresponding epitome model for the first 15 down-scaled frames and compare the up-scaling performance with bicubic interpolation (BIC) and two other methods including a neighbor-embedding based single-image super-resolution (SISR) [15] and a dictionary learning based sparse representations super-resolution (SPARS) [19]. For SISR and SPARS, we employ 20 iterations of back-projection (IBP) [20] as post-processing. For our algorithm, we set $n = 4$ for *Foreman*, and $n = 8$ for *Barbara* and *City*. $K = 2n$.

Fig. 4 illustrates test images and the obtained epitomes with our method and the method in [9] where both epitomes (having different generative models) reconstruct input image at the same quality. One can clearly observe that the proposed method leads to a more compact representation thanks to our multi-patch scheme and optimization over whole image. Note here that we obtain $\mathbf{I}_d$ using lanczos-3 filters in each epitome model $(\mathbf{E}, \mathbf{I}_d)$.

Fig. 5 further draws up-scaling performance of the epitome models $(\mathbf{E}, \mathbf{E}_d)$ obtained from the first frames of *Foreman* and *City*, and

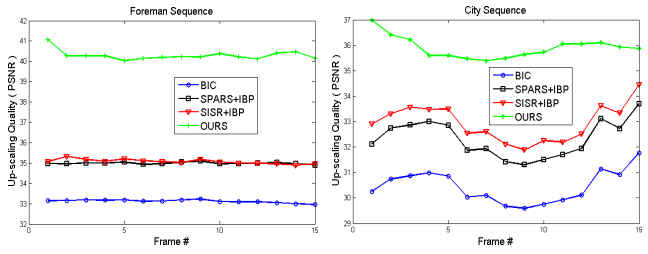

Fig. 5: Video up-scaling performance of epitome models of *Foreman* and *City*.

then applied to first 15 down-scaled frames of these sequences. An obvious and stable reconstruction quality improvement can be observed throughout frames when compared to other methods.

Fig. 6 demonstrates some visual reconstruction results. On the top, we compare our result with SPARS for the *Barbara* image. Bottom four close-up images are extracted from the reconstruction error (in false color) of the 7-th frame of the *City* sequence with our method (using the epitome model of the first frame) in comparison to other methods. One can observe that highly-complex and textured areas have been better reconstructed with less aliasing artifacts using our algorithm (better viewed in the electronic version of the paper).

4. CONCLUSION

We presented a novel epitomic image representation scheme aligned with the manifold sampling assumption through neighbor-embedding. We end up a generic signal representations with increased optimality, compactness, and stability. Because of the page limitation, we demonstrated the performance of our method for only image and video up-scaling. Future work includes extending this framework to image and video compression, segmentation, and classification.

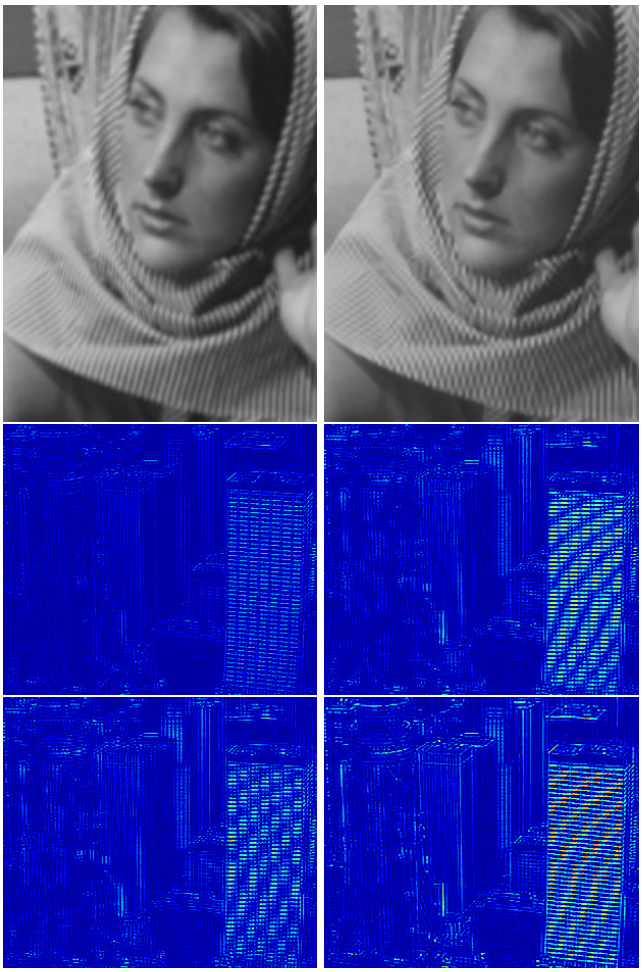

Fig. 6: (Left-to-right top-to-bottom) Reconstruction of *Barbara* with our method versus SPARS; Reconstruction error of 7-th frame of *City* with our method, SPARS with IBP; SISR with IBP, and BIC.

5. REFERENCES

[1] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recogn.*, 2008, pp. 1–8.

[2] L.-Y. Wei, J. Han, K. Zhou, H. Bao, B. Guo, and H.-Y. Shum, "Inverse texture synthesis," in *ACM Trans. Graphics (SIGGRAPH)*, 2008, pp. 1–9.

[3] N. Jojic, B. J. Frey, and A. Kannan, "Epitomic analysis of appearance and shape," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2003, pp. 34–41.

[4] V. Cheung, B. J. Frey, and N. Jojic, "Video epitomes," in *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recogn.*, 2005, pp. 42–49.

[5] A. Kannan, J. Winn, and C. Rother, "Clustering appearance and shape by learning jigsaws," in *Advances in Neural Information Process. Syst.*, 2007, pp. 657–664.

[6] H. Wang, Y. Wexler, E. Ofek, and H. Hoppe, "Factoring repeated content within and among images," in *ACM Trans. Graphics (SIGGRAPH)*, 2008.

[7] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2009, pp. 349–356.

[8] T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2013, pp. 945–952.

[9] S. Cherigui, C. Guillemot, D. Thoreau, P. Guillotel, and P. Perez, "Epitome-based image compression using translational sub-pel mapping," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2011, pp. 1–6.

[10] S. Cherigui, M. Alain, C. Guillemot, D. Thoreau, and P. Guillotel, "Epitome inpainting with in-loop residue coding for image compression," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 5581–5585.

[11] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[12] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[13] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. National Academy Sci.*, vol. 100, no. 10, pp. 5591–5596, 2003.

[14] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recogn.*, 2004, pp. 275–282.

[15] M. Turkan, D. Thoreau, and P. Guillotel, "Self-content super-resolution for ultra-HD up-sampling," in *Proc. European Conf. Visual Media Prod.*, 2012, pp. 49–58.

[16] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–10, 2010.

[17] M. Turkan, D. Thoreau, and P. Guillotel, "Optimized neighbor embeddings for single-image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 645–649.

[18] M. Turkan, D. Thoreau, and P. Guillotel, "Iterated neighbor-embeddings for image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 3887–3891.

[19] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, pp. 2861–2873, 2010.

[20] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Vis. Comm. Image Repres.*, vol. 4, pp. 324–335, 1993.